



TUGAS AKHIR SM 1332

KLASIFIKASI KANKER PARU DENGAN MENGGUNAKAN ALGORITMA *CLASSIFY BY SEQUENCE (CBS)*

I PUTU EDDY SUKADIANA
NRP 1206 100 711

Dosen Pembimbing
Dr. Imam Mukhlash, S.Si, M.T

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



FINAL PROJECT SM 1332

LUNG CANCER CLASSIFICATION USING CLASSIFY BY SEQUENCE (CBS) ALGORITHM

I PUTU EDDY SUKADIANA
NRP 1206 100 711

Supervisor
Dr. Imam Mukhlash, S.Si, M.T

Department Of Mathematics
Faculty Of Mathematics and Sciences
Sepuluh Nopember Institute Of Technology
Surabaya 2015

LEMBAR PENGESAHAN

KLASIFIKASI KANKER PARU DENGAN MENGUNAKAN ALGORITMA *CLASSIFY BY SEQUENCE (CBS)*

LUNG CANCER CLASSIFICATION USING CLASSIFY BY SEQUENCE (CBS) ALGORITHM

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
Pada Bidang Minat Studi Ilmu Komputer
Program Studi S-1 Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

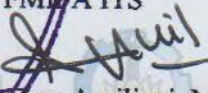
I PUTU EDDY SUKADIANA

NRP. 1206 100 711

Menyetujui,
Dosen Pembimbing,


Dr. Imam Mukhlash, S.Si, MT
NIP. 197006231 199403 1 003

Mengetahui,
Ketua Jurusan Matematika
FMIPA ITS


Prof. Dr. Erna Apriliani, M.Si

NIP. 19660414 199102 2 001

Surabaya, 27 Januari 2015



KLASIFIKASI KANKER PARU DENGAN MENGUNAKAN ALGORITMA *CLASSIFY BY SEQUENCE (CBS)*

Nama : I Putu Eddy Sukadiana
NRP : 1206 100 711
Jurusan : Matematika FMIPA – ITS
Dosen Pembimbing : Dr. Imam Mukhlash, S.Si, MT

Abstrak

Kanker paru-paru merupakan penyakit dengan ciri khas adanya pertumbuhan sel yang tidak terkontrol pada jaringan paru-paru. Salah satu pendekatan (metode) untuk mengidentifikasi kanker paru adalah pendekatan data mining. Beberapa teknik data mining telah banyak dikembangkan seperti teknik asosiasi, clustering, sequence pattern dan klasifikasi. Berkembang pula metode klasifikasi dengan penggabungan sequential pattern mining dan induksi probabilitas yang dikenal sebagai algoritma Classify-by-Sequence (CBS). Belum ada kajian yang dilakukan mengenai data mining untuk data rekam medis dengan memanfaatkan algoritma Classify-by-Sequence (CBS).

Pada penelitian ini akan dilakukan klasifikasi data rekam medis pada Rumah Sakit Umum Daerah (RSUD) di Denpasar, Bali, yang diolah dengan menggunakan Matlab R2009a. Diharapkan pada penelitian ini akan menunjukkan hasil akurasi yang tinggi meskipun hanya menggunakan dataset yang kecil.

Kata Kunci: Data mining, Classify by Sequence, CBS, Rekam medis, Kanker paru.

LUNG CANCER CLASSIFICATION USING CLASSIFY BY SEQUENCE (CBS) ALGORITHM

Name : I Putu Eddy Sukadiana
NRP : 1206 100 711
Department : Mathematics FMIPA – ITS
Supervisor : Dr. Imam Mukhlash, S.Si, MT

Abstract

Lungs cancer is a diseases that has a characteristic of abnormal sel growth in the lungs. One of the method use to identify cancer is the data mining. A few technique for data mining had been developed, such as association, clustering, sequence pattern and classification. Other method for classification also been developed, wich combine the sequential pattern mining with probability induction known as Classify by Sequence. There aren't any research that have been done about data mining for medical record that using Classify by Sequence algorithm.

This research will classify the patient medical records in regional public hospital in Denpasar, Bali, that will be process in Matlab R2009a. hopefully on this research will show high accuracy value, despite using small dataset.

Keywords: Data Mining, Classify by Sequence, Medical Record, Lung Cancer.

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa karena atas karunia-Nya penulis mampu menyelesaikan Tugas Akhir yang berjudul **“Klasifikasi Kanker Paru dengan Menggunakan Algoritma Classify by Sequence (CBS)”** sebagai syarat kelulusan dalam menempuh program S-1 Jurusan Matematika FMIPA ITS Surabaya.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kata sempurna. Untuk itu penulis mengharap saran dan kritik yang membangun. Sebagai penutup penulis mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak yang memberi bantuan dan dukungan dalam penyusunan Tugas Akhir ini, yang terhormat :

1. Ibu Dr. Erna Apriliani, M.Si selaku Ketua Jurusan Matematika FMIPA ITS.
2. Bapak Dr. Imam Mukhlash, S.Si, MT selaku Dosen Pembimbing Tugas Akhir yang telah banyak membantu, membimbing dan memberikan motivasi kepada penulis dalam penyusunan laporan Tugas Akhir ini.
3. Bapak Dr. Chairul Imron, MI.Komp selaku Koordinator Tugas Akhir Jurusan Matematika FMIPA ITS.
4. Bapak Drs. Soetrisno, MI.Komp. , selaku dosen wali penulis yang telah memberikan motivasi.
5. Bapak dan ibu Dosen serta seluruh staf Tata Usaha dan Laboratorium Jurusan Matematika FMIPA ITS yang tidak dapat penulis sebutkan satu persatu.

Akhir kata penulis mohon maaf sebesar-besarnya apabila ada kesalahan yang diperbuat, baik sengaja maupun tidak. Semoga tugas akhir bermanfaat bagi pembaca khususnya dan rekan-rekan mahasiswa lainnya.

Surabaya, Januari 2015
Penulis

Special thank's to

- Tuhan Yang Maha Esa yang senantiasa memberi kesehatan, kelancaran dan kemudahan dalam pengerjaan Tugas Akhir ini.
- Alm. Mama yang telah melahirkan saya dan percaya bahwa saya mampu serta selalu menyemangati saya.
- Papa dan adik saya, Made Evi Diana Putri Agneswary serta ponakan ku tercinta Gede Nathan Kameswara yang selalu menyemangati saya walau mereka berada di pulau seberang (Bali)
- Pacar saya tercinta, Fatimah Nandita Anggraeni (Dita) yang senantiasa menemani, memberi semangat, menghibur dan selalu membantu saya dalam pengerjaan serta percaya bahwa saya mampu menyelesaikan Tugas Akhir ini.
- Muhammad Iqbal, yang selalu memberi dukungan moral,semangat serta membantu saya jika program yang saya buat terjadi error.
- Charis, Galuh, Dito, Prima Adi, Boneng dan teman-teman warkop saya yang selalu menemani dang menghibur saya dikala penat dan memberi semangat.
- Cak Sholikin dan Edi Sigit yang setia membuatkan kopi yang nikmat untuk menemani saya begadang mengerjakan TA ini.
- Rifdy, adik angkatan ku yang selalu memberi semangat dan yakin bahwa saya pasti bisa.

- Inkboz, Yudha, dan Febrian teman kos saya yang tak letih-letih membangunkan saya.
- Romi, Andi, Yossi dan Solikh, teman-teman seperjuangan wisuda 111 yang saling memberi semangat dan saling berbagi informasi dan saling membantu selama pengerjaan TA ini.
- Teman-teman angkatan saya yang selalu menyemangati saya bahwa saya pasti bisa.
- Teman-teman Himpunan Mahasiswa Matematika ITS (HIMATIKA ITS) yang juga memberi dukungan moral dan semangat kepada saya.

“Halaman ini sengaja dikosongkan”.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	vi
ABSTRAK	vii
<i>ABSTRACT</i>	ix
KATA PENGANTAR	xi
DAFTAR ISI	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
DAFTAR NOTASI	xxiii

BAB I PENDAHULUAN

1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	3
1.6 Sistematika Penulisan	4

BAB II TINJAUAN PUSTAKA

2.1 Data Mining	5
2.2 Teknik dalam Data Mining	6
2.3 Algoritma CBA	7
2.4 Algoritma CBC	8
2.5 Algoritma CBS	10
2.6 Kanker	16
2.7 Kanker Paru	16

BAB III METODE PENELITIAN

3.1 Identifikasi Masalah	21
3.2 Studi Literatur	21
3.3 Pengumpulan Data	21
3.4 Pengolahan Data	21

3.4.1 Analisis Metode dan Perancangan Perangkat Lunak	22
3.4.2 Pengujian dan Evaluasi Perangkat Lunak ...	22
3.5 Penarikan Kesimpulan dan Saran	22
3.6 Penyusunan Laporan Tugas Akhir	22
BAB IV ANALISA DAN PERANCANGAN PERANGKAT LUNAK	
4.1 Analisa Sistem	25
4.2 Pengambilan dan Pengolahan Data	26
4.2.1 Pengambilan Data.....	26
4.2.2 Pengolahan Data.....	27
4.2.3 Preprocessing Data.....	27
4.3 Penemuan Rule Menggunakan Algoritma CBS	28
4.3.1 Menemukan CSP	28
4.3.2 Menemukan Frequent Itemset	28
4.3.3 Membuat CSP Set.....	28
4.3.4 Membuat Classifier.....	29
4.4 Menentukan Nilai Akurasi	29
4.5 Perancangan Antarmuka.....	29
4.6 Implementasi Antarmuka	30
4.7 Perancangan Kode Program	31
BAB V PEMBAHASAN DAN PENGUJIAN	
5.1 Ruang Lingkup Uji Coba.....	39
5.2 Uji Coba	40
BAB VI PENUTUP	
6.1 Kesimpulan	51
6.2 Saran	51
DAFTAR PUSTAKA	53
LAMPIRAN	
A. Cara Menjalankan Perangkat Lunak	55
B. Dataset yang Digunakan	55

C. Listing Program	75
D. Hasil Running	97
E. Tabel Hasil Running Beserta Waktunya.....	106
F. Tabel Data Untuk Running Program	106

BIODATA PENULIS

"Halaman ini sengaja dikosongkan"

DAFTAR TABEL

Tabel 5.1	Nilai akurasi, cakupan dan waktu proses tiap <i>min_sup</i>	49
Tabel B	Dataset.....	55
Tabel E	Hasil Running Beserta Waktunya.....	106
Tabel F	Data yang Digunakan untuk Running Program (Setelah dilakukan modifikasi).....	106

"Halaman ini sengaja dikosongkan"

DAFTAR GAMBAR

Gambar 2.1	Tahapan Data Mining	6
Gambar 2.2	Tahapan Algoritma CBS	11
Gambar 2.3	Algoritma CSP-Miner	12
Gambar 2.4	Perpanjangan dari Pohon CSP.....	13
Gambar 2.5	Algoritma Classifier-Builder.....	14
Gambar 2.6	Perbandingan paru-paru perokok dan bukan perokok.....	17
Gambar 3.1	Diagram Alir Penelitian	23
Gambar 4.1	Perancangan Antarmuka.....	31
Gambar 4.2	Kode Program Memasukkan Data	32
Gambar 4.3A	Kode Program <i>Preprocessing</i>	33
Gambar 4.3B	Lanjutan Kode Program <i>Preprocessing</i> 1	34
Gambar 4.4	Kode Program Untuk Mencari Nilai <i>Frequent Itemset</i>	35
Gambar 4.5	Kode Program untuk Membangun <i>Classifier</i>	36
Gambar 4.6	Kode Program untuk Mencari Akurasi	37
Gambar 5.1	Hasil Uji Coba untuk Tampilan Awal	40
Gambar 5.2	Hasil Uji Coba untuk nilai <i>Min_sup</i> 0,9.....	41
Gambar 5.3	Rule pada saat nilai <i>Min_sup</i> 0,9.....	41
Gambar 5.4	Hasil Uji Coba untuk nilai <i>Min_sup</i> 0,8.....	42
Gambar 5.5	Rule pada saat nilai <i>Min_sup</i> 0,8	42
Gambar 5.6	Hasil Uji Coba untuk nilai <i>Min_sup</i> 0,7.....	43
Gambar 5.7	Rule pada saat nilai <i>Min_sup</i> 0,7.....	43
Gambar 5.8	Hasil Uji Coba untuk nilai <i>Min_sup</i> 0,4.....	44
Gambar 5.9A	Rule pada saat nilai <i>Min_sup</i> 0,4.....	45
Gambar 5.9B	Rule pada saat nilai <i>Min_sup</i> 0,4.....	45
Gambar 5.10	Hasil Uji Coba untuk nilai <i>Min_sup</i> 0,1	46
Gambar 5.11A	Rule pada saat nilai <i>Min_sup</i> 0,1	46
Gambar 5.11B	Rule pada saat nilai <i>Min_sup</i> 0,1	47
Gambar 5.11C	Rule pada saat nilai <i>Min_sup</i> 0,1	47
Gambar 5.11D	Rule pada saat nilai <i>Min_sup</i> 0,1	48
Gambar D1	Min Sup 0,9	97

Gambar D2	Min Sup 0,8	98
Gambar D3	Min Sup 0,7	99
Gambar D4	Min Sup 0,6	100
Gambar D5	Min Sup 0,5	101
Gambar D6	Min Sup 0,4	102
Gambar D7	Min Sup 0,3	103
Gambar D8	Min Sup 0,2	104
Gambar D9	Min Sup 0,1	105

DAFTAR NOTASI

min_support	= Nilai keyakinan minimum yang digunakan sebagai treshold untuk proses <i>pruning</i> .
CSP_set	= Rule sementara yang didapat dari pengolahan data yang diperoleh oleh Algoritma CBS.
Jum rule	= Jumlah rule yang ditemukan.
Dt	= Matriks yang memuat atribut dan kelas (data temporer)
Dc1	= variabel untuk menyimpan data dari Dt.

"Halaman ini sengaja dikosongkan"

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker paru-paru merupakan penyakit dengan ciri khas adanya pertumbuhan sel yang tidak terkontrol pada jaringan paru-paru. Bila tidak dirawat pertumbuhan sel ini dapat menyebar keluar dari paru-paru melalui suatu proses yang disebut metastasis ke jaringan yang terdekat atau bagian tubuh yang lainnya. Kanker paru dapat dilihat melalui foto rontgen dada dan tomografi komputer (*CT scan*). Diagnosis dapat dipastikan dengan biopsi (pengambilan jaringan tubuh untuk pemeriksaan laboratorium) yang biasanya dilakukan melalui prosedur bronkoskopi atau dipandu dengan CT. Kanker paru sebagai satu dari tiga penyakit yang paling mematikan di dunia memiliki perkembangan yang sangat pesat. Beberapa faktor yang mempengaruhi pesatnya perkembangan penyakit kanker paru di dunia antara lain adalah paparan dalam jangka waktu yang lama terhadap asap tembakau, faktor genetik, gas radon dan polusi udara.[1]

Salah satu pendekatan (metode) untuk mengidentifikasi kanker paru adalah pendekatan *data mining*. *Data mining* adalah salah satu cabang ilmu di bidang komputer, dimana perkembangannya sangat pesat. *Data mining* telah banyak diaplikasikan ke dalam berbagai bidang seperti ilmu pengetahuan, teknik dan bisnis. Tujuan utama dari *data mining* adalah untuk menemukan informasi penting dalam *database* yang sangat besar. Beberapa teknik *data mining* telah banyak dikembangkan seperti teknik asosiasi, *clustering*, *sequence pattern* dan klasifikasi. Waktu merupakan salah satu hal yang penting dalam data. Banyak data dalam kehidupan nyata yang menggambarkan hal atau keadaan beberapa obyek pada waktu tertentu yang disebut data temporal [2].

Kemudian berkembang suatu metode klasifikasi dengan penggabungan *sequential pattern mining* dan induksi probabilitas

yang dikenal sebagai algoritma *Classify-by-Sequence (CBS)*. Algoritma *CBS* menurut Tseng dan Lee dapat memberikan:

1. Performansi yang baik pada klasifikasi *dataset* temporal,
2. Informasi klasifikasi kepada *user* pada saat bersamaan,
3. Efisiensi eksekusi tinggi menggunakan induksi probabilitas dan
4. Hasil klasifikasi tersedia.

Selain algoritma *CBS*, berkembang pula algoritma *feature-mining* yang mengambil sekumpulan sifat dan menambahkannya dengan sifat makro untuk menguji kemunculan dari nilai kombinasi yang sesungguhnya. Selain itu algoritma ini juga memiliki keuntungan yaitu proses pembelajarannya dapat melampaui bias mula-mula dan hipotesa agar dapat lebih dekat pada tujuan konsep yang sebenarnya[3].

Terdapat juga penelitian yang menggunakan metode *data mining* yang lain seperti menggunakan K-mean *clustering* untuk kanker paru, tetapi memiliki tingkat akurasi rendah. Sehingga diperlukan perlakuan tambahan dengan program khusus untuk meningkatkan akurasinya[4]. Sama halnya dengan metode klasifikasi SVMs yang memiliki tingkat akurasi yang lemah karena banyaknya data yang salah dalam proses klasifikasinya[5].

Kajian mengenai *data mining* untuk data rekam medis belum banyak dilakukan dengan memanfaatkan algoritma *Classify-by-Sequence (CBS)*. Elemen yang dibutuhkan dalam data rekam medis yaitu analisis klinis, hasil rontgen toraks dan patologi anatomis. Berdasarkan kajian-kajian tersebut, maka pada Tugas Akhir ini akan diteliti tentang klasifikasi berdasarkan *Classify-by-Sequence (CBS)* menggunakan data yang terdapat pada rekam medis pasien.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka permasalahan yang akan diselesaikan dalam Tugas Akhir ini adalah:

- a. Bagaimana pendekatan algoritma *Classify-by-Sequence (CBS)* untuk membantu klasifikasi dalam data rekam medis pasien;
- b. Bagaimana mengaplikasikan algoritma *Classify-by-Sequence (CBS)* untuk membantu klasifikasi dalam data rekam medis pasien;

1.3 Batasan Masalah

Dalam buku Tugas Akhir ini, permasalahan yang akan diselesaikan dibatasi ruang lingkupnya yaitu sebagai berikut:

- a. Data yang digunakan adalah data primer
- b. Data yang digunakan adalah data rekam medis pasien Rumah Sakit Umum Daerah Wangaya, Denpasar, Bali

1.4 Tujuan

Berdasarkan latar belakang dan permasalahan yang dikemukakan maka tujuan dari penulisan tugas akhir ini adalah membangun perangkat lunak untuk klasifikasi data berdasarkan rekam medis pasien yang memuat algoritma *Classify-by-Sequence (CBS)*.

1.5 Manfaat

Manfaat yang ingin diperoleh dari penyelesaian permasalahan dalam buku Tugas Akhir ini adalah:

- a. Mendapatkan pola *sequence* yang berguna dalam data rekam medis pasien untuk membantu didapatkannya klasifikasi.
- b. Menghasilkan suatu perangkat lunak yang menerapkan algoritma klasifikasi berbasis aturan *sequence* untuk data rekam medis pasien dengan nilai akurasi yang tinggi.

1.6 Sistematika Penulisan

Tugas akhir ini disusun berdasarkan sistematika tulisan sebagai berikut:

BAB I PENDAHULUAN

Berisi tentang latar belakang, rumusan dan batasan masalah serta tujuan dan manfaat penelitian.

BAB II TINJAUAN PUSTAKA

Memaparkan dasar teori yang digunakan penulis dalam mengerjakan tugas akhir.

BAB III METODOLOGI PENELITIAN

Menjelaskan alur kerja dan metode yang digunakan penulis dalam mengerjakan tugas akhir.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini menjelaskan mengenai implementasi dan hasil pengujian implementasi pada setiap proses dalam mengidentifikasi cipherteks dan melakukan pembahasan terhadap hasil yang diperoleh.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan penutup, berisi tentang kesimpulan yang dapat diambil dan saran yang selayaknya dilakukan bila tugas akhir ini dilanjutkan.

BAB II

TINJAUAN PUSTAKA

Dalam pengerjaan tugas akhir ini diperlukan tinjauan pustaka yang menjadi dasar teori serta pendukung untuk menyelesaikan permasalahan yang diteliti.

2.1 *Data mining*

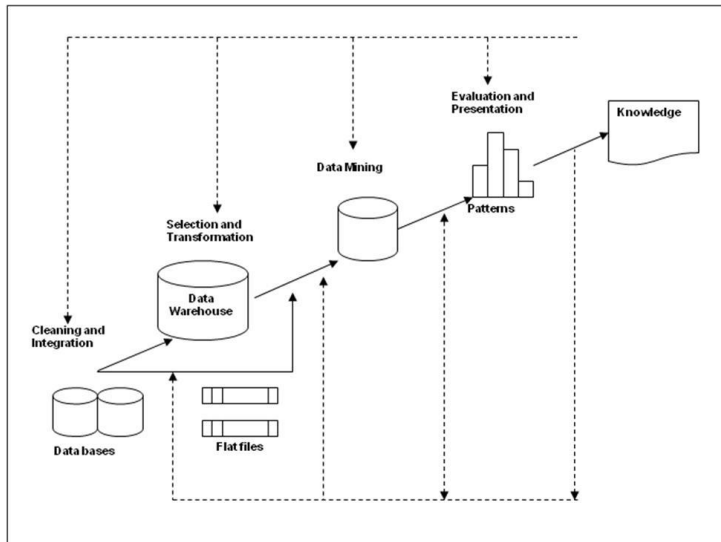
Data mining secara sederhana dapat diartikan sebagai ekstraksi atau “menambang” pengetahuan dari data dalam jumlah besar, meskipun *data mining* seharusnya lebih tepat dinamai “penggalian pengetahuan dari data” yang terlalu panjang pada penamaannya. Namun “penggalian pengetahuan” sebagai kata yang lebih singkat tidak dapat mewakili penekanan pada data besar. Meski demikian *mining* adalah istilah yang jelas untuk menggambarkan suatu proses untuk menemukan data kecil yang berharga dari data yang besar[6].

Dalam *data mining*, terdapat tahap-tahap yang harus dilalui dalam proses untuk mendapatkan pengetahuan dan melibatkan rangkaian yang berulang dari tahap-tahap di bawah ini:

1. *Data cleaning*
Proses untuk menghilangkan noise dan data yang tidak konsisten.
2. *Data integration*
Proses dimana beberapa sumber data yang telah diperoleh digabungkan menjadi satu.
3. *Data selection*
Proses dimana data yang relevan untuk proses analisis diambil dari database.
4. *Data transformation*
Proses dimana data diubah atau digabungkan menjadi data yang sesuai untuk diolah dengan menggunakan operasi agregasi.
5. *Data mining*

Suatu proses penting dimana metode pintar diterapkan untuk ekstraksi pola data.

6. *Pattern evaluation*
7. Proses untuk mengenali pola yang benar-benar menarik yang mewakili pengetahuan yang ingin didapat.
8. *Knowledge presentation*
Proses dimana visualisasi dan gambaran suatu pengetahuan digunakan untuk menampilkan pengetahuan yang didapat pada pengguna[6].



Gambar 2.1 Tahapan *Data mining*

2.2 Teknik dalam *Data Mining*. [6]

Untuk mendapatkan data-data yang diinginkan, *data mining* memiliki beberapa teknik yang populer dan sering digunakan. Teknik-teknik populer tersebut antara lain :

1. *Association rule mining*

Association rule mining adalah teknik *data mining* untuk menemukan aturan assosiatif antara satu kombinasi item dengan kombinasi item lainnya.

2. **Klasifikasi**

Klasifikasi merupakan tugas menentukan objek pada salah satu kategori yang telah diketahui sebelumnya. Klasifikasi juga didefinisikan sebagai berikut :

Klasifikasi merupakan tugas pembelajaran fungsi target f yang memetakan setiap himpunan atribut x pada label *class* y yang telah didefinisikan sebelumnya. Fungsi target disebut juga sebagai model klasifikasi.

3. **Clustering**

Clustering adalah pengelompokan data tanpa pengelompokan kelas tertentu dan dapat dipakai untuk memberikan label pada kelas data yang belum diketahui.

4. **Sequential Pattern Mining**

Sequential pattern mining merupakan salah satu teknik klasifikasi data temporal. Tujuannya untuk menemukan *frequent sub sequences* sebagai pola dalam *sequence database*. Menurut Han, J., dkk *sequential pattern mining* didefinisikan sebagai “diberikan *sequence database* dan ambang batas *min support*, *sequential pattern mining* digunakan untuk menemukan *complete set* (himpunan lengkap) dari *sequential patterns* pada *database*[7].

2.3 Algoritma CBA

Salah satu algoritma paling awal dan paling sederhana untuk klasifikasi asosiatif adalah *CBA* (*Classification-based-Association*). *CBA* menggunakan pendekatan iteratif pada *item set mining* yang hampir sama pada algoritma Apriori, dimana beberapa langkah dibuat melalui suatu data dan item set awal

digunakan untuk menghasilkan dan mengetes *item set* yang lebih panjang. Algoritma Apriori sendiri merupakan algoritma yang paling banyak digunakan untuk menemukan *frequent itemset* dan aturan asosiasi. Konsep utama dari algoritma Apriori adalah sebagai berikut :

1. Setiap subset dari *frequent itemset* adalah *frequent itemset*.
2. Himpunan *itemset* dengan panjang k disebut C_k .
3. Himpunan *itemset* yang memenuhi batasan minimum support disebut sebagai L_k , L_k adalah kandidat himpunan yang digunakan untuk tahap selanjutnya.
4. C_{k+1} dibangkitkan dengan L_k dan dirinya sendiri. *Itemset-itemset* yang memenuhi kriteria masing-masing memiliki lebih satu elemen dari *itemset* sebelumnya.

L_k kemudian dihasilkan dengan menghilangkan elemen-elemen yang tidak memenuhi aturan *minimum support*. Sebab, kandidat *sequence* yang dibangkitkan dimulai dengan ukuran *sequence* terkecil secara bertahap ukuran *sequence* meningkat yang disebut dengan pendekatan *breadth first search* [8].

CBA menggunakan metode heuristik untuk membangun classifier-nya, dimana beberapa aturan disusun dari awal hingga akhir berdasarkan tingkat keyakinan dan faktor pendukungnya. Jika suatu aturan memiliki awal yang sama, maka aturan dengan tingkat keyakinan yang lebih tinggi dipilih untuk mewakili set tersebut[6].

2.4 Algoritma CBC

CBC (*Classification-based-Clustering*) adalah suatu model yang garis besarnya hampir mirip dengan metode *decision tree* (pohon keputusan) tetapi keunggulan utamanya adalah selalu menyediakan aturan pengambilan keputusan yang tepat, berbeda dengan *decision tree* yang terkadang gagal untuk menyediakan aturan yang dapat dipakai oleh peneliti. Seperti semua metode

klasifikasi, metode ini juga menggunakan metodologi *training* dan pengujian set.

Algoritma CBC memiliki 6 tahap proses antara lain :

1. Memilih atribut target.
Karena CBC merupakan suatu model klasifikasi, atribut target sangatlah penting. Atribut target haruslah berupa data kategorial.
2. Menjalankan algoritma clustering pada dataset.
Algoritma clustering yang digunakan pada tahap ini dapat menggunakan algoritma apapun. Jika algoritma yang digunakan membutuhkan parameter dari peneliti, disarankan untuk menjaankan algoritma beberapa kali menggunakan parameter yang berbeda untuk mendapat parameter yang paling sesuai.
3. Menghitung distribusi atribut target pada setiap cluster.
Setiap cluster mengandung bagian dari entitas pada dataset. Kelompok entitas ini memiliki distribusi masing-masing pada atribut target. Jika atribut target berupa data biner, maka distribusinya disebut laju reaksi dari cluster tersebut.
4. Menentukan *threshold*
Distribusi yang telah dihitung dari atribut target pada setiap cluster sebenarnya adalah kemungkinan dari entitas dengan atribut yang sama memiliki semua nilai yang mungkin dari atribut target.
5. Menyempurnakan hasil
Karena model clustering dirancang tanpa menggunakan atribut target dan tidak memiliki proses validasi maka kemungkinan menjadi kurang mampu untuk model konvensional. Untuk mengatasi masalah ini maka

algoritma clustering perlu untuk disempurnakan dengan memberikan bobot tambahan untuk atribut target.

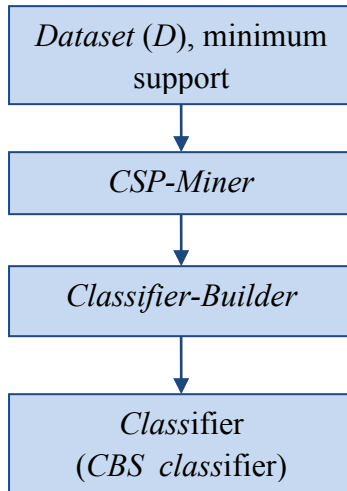
6. Menguji hasil

Hasil dari algoritma clustering dijalankan pada dataset baru dan klasifikasikan entitas berdasarkan pada algoritma clustering tersebut.[3]

2.5 Algoritma CBS

Misalkan terdapat *database D* sangat besar yang menyimpan sejumlah *data event*, dimana kejadian data terdiri atas *sequence* data temporal. Setiap *record* data berkaitan dengan sebuah *class*. Misal D_i merupakan semua kejadian data temporal yang termasuk dalam sebuah *class*. Selanjutnya, *database* $D = \{D_1, D_2, \dots, D_N\}$ dan diasumsikan terdapat N -class dalam *database*. Untuk setiap *class*, *dataset* D_i terdiri atas sejumlah kejadian data temporal $I_j = \langle a_{j1}, a_{j2}, \dots, a_{ji} \rangle$ dengan a_{jk} merupakan nilai waktu ke- k dari kejadian I_j . [9]

Algoritma *CBS* terdiri atas dua tahap, tahap pertama untuk *Mining classifiabel sequence* dan tahap kedua merupakan barisan yang telah ditentukan pada tahap pertama digunakan untuk membangun klasifier. Secara umum ditunjukkan dalam **Gambar 2.2.** berikut :



Gambar 2.2 Tahapan Algoritma CBS.

Input : dataset (D),
Minimum *support* (*min_sup*)
Output : CBS_classifier
Langkah : CBS (D, *min_sup*)
 CSP_feature_set = CSP-miner(D, *min_sup*)
 CBS_classifier = classifier-builder(CSP_feature set)
 output

Algoritma CBS dikembangkan oleh V.S. Tseng dan C.H. Lee. Berdasarkan penelitian yang mereka lakukan CBS memiliki 2 tahap, yang pertama adalah untuk mencari urutan yang dapat diklasifikasikan, dimana urutan ini diambil dari setiap grup kelas. Dan yang kedua adalah menggunakan urutan yang ditemukan pada tahap pertama untuk membuat classifiernya. Konsep dari CBS adalah untuk meng-ekstrak *time sequence feature* dari setiap kelompok kelas untuk menciptakan aturan klasifikasi, dengan cara ini diharapkan classifiernya menjadi lebih akurat. Kedua tahap tersebut adalah:[6]

1. Tahap pertama : Algoritma CSP – Miner

```

Algorithm CSP- Miner
Input : a temporal dataset (D),
          minimum support (min_sup)
Output : all CSPs of each class ( {CSP_setc | c ∈ all_class_of(D)} )
Procedure CSP- Miner( D, min_sup )
1.  for each class c ∈ all_class_of(D) do
2.    let Dc be the dataset consists of all data in D belong to class c
3.    F1 = frequent item set of Dc;
4.    CSP_setc = F1;
5.    let T be a prefix tree build by F1
6.    k=2;
7.    while T.root is exists
8.      set T' be the tree generated form extension of T
9.      for each temporal data s in Dc do
10.        trace T' with s and add one in leaves node it reaches.
11.      end for
12.      T = T' after non-frequent subtree pruning
13.      Fk = all sequence in T
14.      CSP_setc = CSP_setc ∪ Fk
14.     k=k+1
15.    end while
16.  end for
17. output {CSP_setc | c ∈ all_class_of(D)}

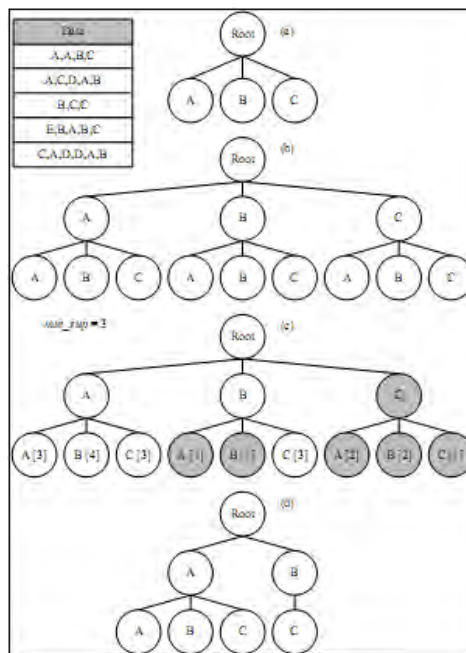
```

Gambar 2.3 Algoritma CSP - Miner

Pada tahap ini akan dicari semua pola urutan yang dapat diklasifikasikan (CSP) untuk klasifikasi data temporal. Pertama akan dicari item yang sering muncul dari keseluruhan *dataset* sepanjang satu kali urutan yang sering muncul. Lalu akan dibuat akar yang memuat item-item ini sebagai daunnya. Kemudian dengan arsitektur ini CBS akan dapat menemukan semua urutan yang memungkinkan. Selanjutnya perluasan dari pohon akan digunakan untuk menghasilkan semua kandidat yang memungkinkan.

Pada pohon ini semua elemen akan dikumpulkan ke daunnya, untuk mendapatkan semua elemen dengan urutan k

dan dilanjutkan dengan urutan $k+1$. Setelah semua pohon terbentuk, lalu dihitung untuk keseluruhan *dataset* yang digunakan untuk proses *pruning* dan bagian yang tidak sering muncul. Setiap *transaction* harus berupa rantai item, sehingga prefixnya dapat ditelusuri dengan mudah. Akan dihitung setiap waktu singgah. Setelah dihitung, hilangkan semua sub pohon yang tidak mengandung daun yang memenuhi CSP, proses ini akan membuat semua cabang pohon memiliki kedalaman yang sama.



Gambar 2.4 Perpanjangan dari pohon CSP

Gambar di atas menunjukkan jalannya tahap pertama dari algoritma CBS, dari data yang ada kita dapatkan semua item yang sering muncul pada dataset. Lalu CSP_miner akan membangun

pohon awal T . Berarti pohon T dibangun dari semua rangkaian yang sering muncul yang memiliki panjang sama dengan 1. Lalu dibangun semua rangkaian kandidat dengan panjang 2. Sehingga didapat pohon baru T' seperti yang ditunjukkan pada gambar 2.4(b), lalu dilakukan penghitungan data yang sering muncul dengan berpedoman pada dataset. Dengan nilai *min_sup* pada contoh adalah 3, maka, daun yang terdapat pada lapisan terbawah dari pohon T' yang memiliki nilai *min_sup* di bawah 3 akan dieliminasi seperti yang ditunjukkan pada gambar 2.4(c). setelah proses eliminasi, maka hasil akhir dari proses setelah semua sub pohon tanpa daun yang sering muncul ditunjukkan pada gambar 2.4(d).

2. Tahap kedua : Algoritma pembangun Classifier

<p>Algorithm Classifier- Builder</p> <p>Input : the CSP set of each class ($\{CSP_set_x \mid x \in all_class_of(D)\}$)</p> <p>Output : a CBS classifier (T)</p> <p>Procedure Classifier- Builder ($\{CSP_set_x \mid x \in all_class_of(D)\}$)</p> <ol style="list-style-type: none"> 1. let CSP_dup be the CSP set consist of the CSPs which exist in all class 2. for each $x \in all_class_of(D)$ do 3. for each CSP p in CSP_set_x do 4. if ($p \in CSP_dup$) 5. remove p from CSP_set_x 6. end for 7. end for 8. let T be a prefix tree just have a root node 9. for each $x \in all_class_of(D)$ do 10. for each CSP p in CSP_set_x do 11. put p into T and add a class label at the end node 12. end for 13. end for 14. output T

Gambar 2.5 Algoritma Classifier – Builder

Pada tahap ini akan ditunjukkan bagaimana membangun suatu classifier berdasarkan semua CSP yang ditemukan pada tahap pertama. Pertama – tama semua CSP yang tidak memiliki informasi tentang klasifikasi harus dihilangkan, sehingga tersisa CSP dari masing – masing kelas yang memenuhi saja. Karena setiap CSP memiliki kelasnya tersendiri maka memungkinkan CSP tersebut menjadi kelas pada CSP. Berdasar pada konsep sederhana ini, maka akan dihasilkan suatu metode penilaian tentang bagaimana mengklasifikasikan kejadian temporal yang mengandung banyak CSP.

Akan ditunjukkan bagaimana menggunakan informasi yang didapat dari klasifikasi CSP pada dua dimensi. Pertama anggap seluruh kejadian S memiliki banyak CSP kelas A yang memiliki kemungkinan menjadi kelas A. Kedua atribut dari setiap CSP memiliki informasi penting pada klasifikasi, sehingga setiap *support* dan *confident* dari CSP menjadi penilai yang penting pada masing-masing kelas. Setiap *node* mungkin memiliki lebih dari satu nilai *support*, saat suatu CSP yang diwakilkan oleh sebuah *node* yang dimiliki lebih dari satu kelas. Secara intuisi jika sebuah CSP memiliki lebih banyak nilai *support* maka CSP tersebut mungkin kurang penting untuk klasifikasi. Sehingga tiap-tiap CSP yang kurang tersebut akan dihilangkan. Dengan memperhatikan entrophy dari setiap atribut. Entrophy juga dapat diperlakukan sebagai suatu informasi klasifikasi dan juga sebagai tambahan semakin panjang suatu CSP dari S yang dimiliki oleh kelas A mewakili bahwa S memiliki sifat dari A.

2.6 Kanker

Kanker adalah penyakit akibat pertumbuhan yang tidak normal dari sel-sel jaringan tubuh yang berubah menjadi sel kanker. Dalam perkembangannya sel-sel kanker tersebut dapat menyebar ke bagian tubuh lainnya sehingga dapat menyebabkan kematian. Kanker sering dikenal oleh masyarakat sebagai tumor, padahal tidak semua tumor adalah kanker. Tumor adalah segala benjolan tidak normal atau abnormal. Tumor dibagi dalam 2 golongan yaitu tumor jinak dan tumor ganas. Kanker adalah istilah umum untuk semua jenis tumor ganas.

Kanker dapat menimpa semua orang pada setiap bagian tubuh dan pada semua golongan umur, namun lebih sering menimpa orang yang berusia 40 tahun. Umumnya sebelum kanker meluas atau merusak jaringan di sekitarnya, penderita tidak merasakan adanya keluhan ataupun gejala. Bila sudah ada keluhan atau gejala, biasanya penyakitnya sudah lanjut.

Ada 7 gejala yang perlu diperhatikan dan diperiksa lebih lanjut ke dokter untuk memastikan ada atau tidaknya kanker, yaitu:

1. Waktu buang air besar atau kecil ada perubahan kebiasaan atau gangguan.
2. Alat pencernaan terganggu dan susah menelan.
3. Suara serak atau batuk yang tak sembuh-sembuh.
4. Payudara atau di tempat lain ada benjolan (tumor).
5. Tahi lalat yang berubah sifatnya, menjadi semakin besar dan gatal.
6. Darah atau lendir yang abnormal keluar dari tubuh.
7. Adanya koreng atau borok yang tak mau sembuh-sembuh.[1]

2.7 Kanker paru

Paru-paru merupakan salah satu dari organ penting pada tubuh manusia yang berfungsi untuk membawa oksigen masuk ke tubuh dan mengeluarkan karbondioksida hasil dari sisa pernafasan. Paru-paru juga merupakan organ yang sensitive

terhadap alergi dan paparan radikal bebas seperti polusi udara. Salah satu penyebab utama kanker paru adalah asap rokok. Kanker paru kebanyakan terjadi pada pria yang berusia 40 tahun dan sering berada pada lingkungan asap tembakau. Baik itu perokok aktif maupun perokok pasif.



Gambar 2.6 Perbandingan paru-paru perokok dan bukan perokok .

Kanker paru memiliki beberapa ciri-ciri, antara lain:

1. Sesak nafas.
2. Batuk yang tidak kunjung sembuh.
3. Suara yang menciut saat menarik nafas.
4. Batuk berdarah.
5. Dahak yang dikeluarkan berwarna kuning pekat .
6. Penurunan berat badan secara drastic
7. Terlihat pembengkakan pada bagian leher ataupun wajah.

Namun hal yang paling mencolok adalah batuk berdarah serta dahak yang keluar berwarna kuning pekat dan terkadang menimbulkan bau yang menjadi penanda saat kanker memasuki stadium lanjut.

Biasanya kanker paru baru diketahui pada saat stadium lanjut. Pada stadium ini jika bagian kanker masih dapat diangkat maka pasien masih memiliki peluang sembuh. Namun jika tidak maka pasien akan diarahkan untuk melakukan terapi paliatif, yaitu pengobatan yang hanya diarahkan hanya untuk menghilangkan gejala pasien, membuat pasien merasa lebih baik tanpa harus mengubah perjalanan alami penyakit, mengingat

peluang kesembuhan kanker paru adalah yang paling rendah dibanding dengan kanker yang lainnya.

Untuk tumor bronkial jinak biasanya diangkat melalui pembedahan, karena dapat menyumbat bronki dan semakin lama dapat menjadi ganas. Sekitar 10-15% kanker dapat diangkat melalui pembedahan, tetapi pembedahan tidak selalu membawa kesembuhan.

Sekitar 25-40% penderita tumor yang terisolasi dan tumbuh perlahan memiliki harapan hidup hingga 5 tahun setelah penyakitnya terdiagnosis. penderita harus melakukan pemeriksaan rutin, karena terdapat kemungkinan 6-12% setelah pembedahan kanker paru dapat muncul kembali. Sebelum pembedahan biasanya dokter akan melakukan tes terhadap fungsi paru. Jika hasil dari tes fungsi paru hasilnya tidak seperti yang diharapkan maka kemungkinan pembedahan tidak dilakukan.

Lebih dari 90% kanker paru berawal dari bronki. Bronki adalah saluran udara besar yang masuk ke paru-paru. Kanker ini biasanya disebut karsinoma bronkogenik yang terdiri dari

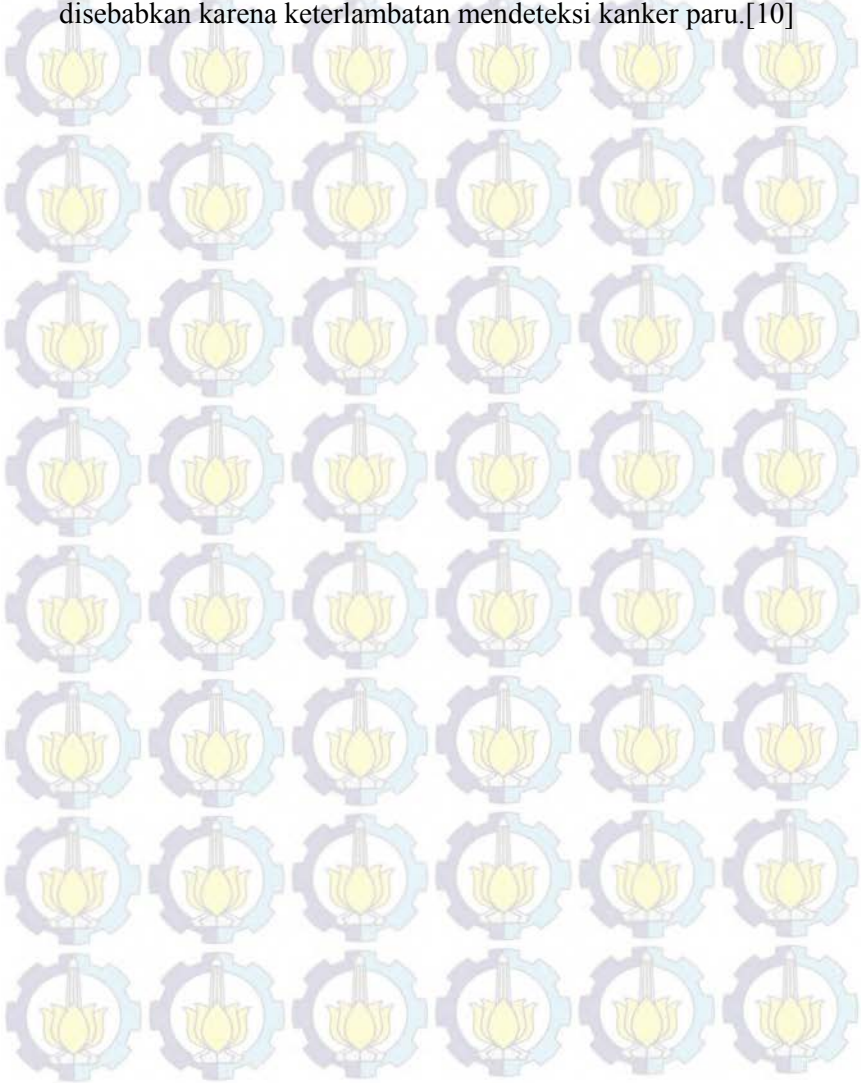
1. Karsinoma sel skuamosa.
2. Karsinoma sel kecil.
3. Karsinoma sel besar.
4. Adenokarsinoma.

Dalam situs Organisasi Kesehatan Dunia (WHO) pada tahun 2012, berikut daftar peringkat penyebab kematian karena kanker:

1. Kanker Paru (1,37 juta kematian)
2. Kanker Perut (736.000 kematian)
3. Kanker Hati (695.000 kematian)
4. Kanker Kolorektal (608.000 kematian)
5. Kanker Payudara (458.000 kematian)
6. Kanker Serviks (275.000 kematian)

Di Indonesia jumlah pasien kanker paru meningkat 20 persen tiap tahunnya. Angka kematian akibat kanker paru 205 per 1 juta orang. Angka kesintasan (survival rate) untuk 5 tahun

kanker paru hanya 13 persen. Jauh lebih kecil dari kanker payudara, prostat dan usus, yang bisa mencapai 80 persen. Hal ini disebabkan karena keterlambatan mendeteksi kanker paru.[10]



"Halaman ini sengaja dikosongkan"

BAB III

METODOLOGI PENELITIAN

Pada bab ini diuraikan metodologi sistem yang digunakan untuk menyelesaikan Tugas Akhir. Pembahasan metodologi penelitian berguna sebagai acuan sehingga penelitian ini dapat berjalan secara sistematis.

3.1 Identifikasi Masalah

Tahap ini bertujuan untuk mendapatkan permasalahan yang dibahas dalam penelitian dan penentuan tujuan penelitian.

3.2 Studi Literatur

Pada tahap ini dilakukan studi yang berkaitan dengan permasalahan yang ada. Studi literatur dilakukan terhadap jurnal-jurnal ilmiah, tugas akhir dan buku-buku yang berhubungan dengan data mining, Algoritma *Classify-by-Sequence* (CBS) dan Kanker paru.

3.3 Pengumpulan Data

Pada tahap ini dilakukan pengambilan data mengenai data rekam medis pasien hingga tahun 2014 untuk proses pengumpulan data. Studi ini dilakukan dengan mengambil data di Rumah Sakit Umum Daerah Wangaya, Denpasar, Bali. Selain itu pada tahap ini juga dilakukan penelitian mengenai klasifikasi dan Aturan *Sequence* pada *Data mining* untuk pendeteksian metode yang akan digunakan. Studi ini dilakukan dengan media buku, jurnal, tesis maupun artikel-artikel di internet.

3.4 Pengolahan Data

Pada tahap ini dilakukan pengolahan terhadap data yang sudah diperoleh. Dimana 80% data digunakan untuk proses *learning* (pembelajaran) untuk menemukan pola data pada algoritma CBS yang digunakan dan 20% datanya digunakan untuk proses *testing* (pengujian).

3.4.1 Analisis Metode dan Perancangan Perangkat Lunak

Tahap ini bertujuan untuk memahami metode yang telah dipilih untuk mengolah data rekam medis pasien. Langkah selanjutnya adalah memahami algoritma dari metode yang dipilih untuk kemudian mengimplementasikannya ke dalam perangkat lunak. Selain itu juga merancang perangkat lunak yang akan dibuat, meliputi perancangan data, proses, dan *interface*.

3.4.2 Pengujian dan Evaluasi Perangkat Lunak

Pada tahap ini dilakukan pengujian terhadap perangkat lunak yang telah dibuat. Pengujian meliputi proses perangkat tersebut bekerja dan hasil yang ditampilkan. Kemudian dilakukan evaluasi untuk menyempurnakan perangkat lunak yang telah dibuat.

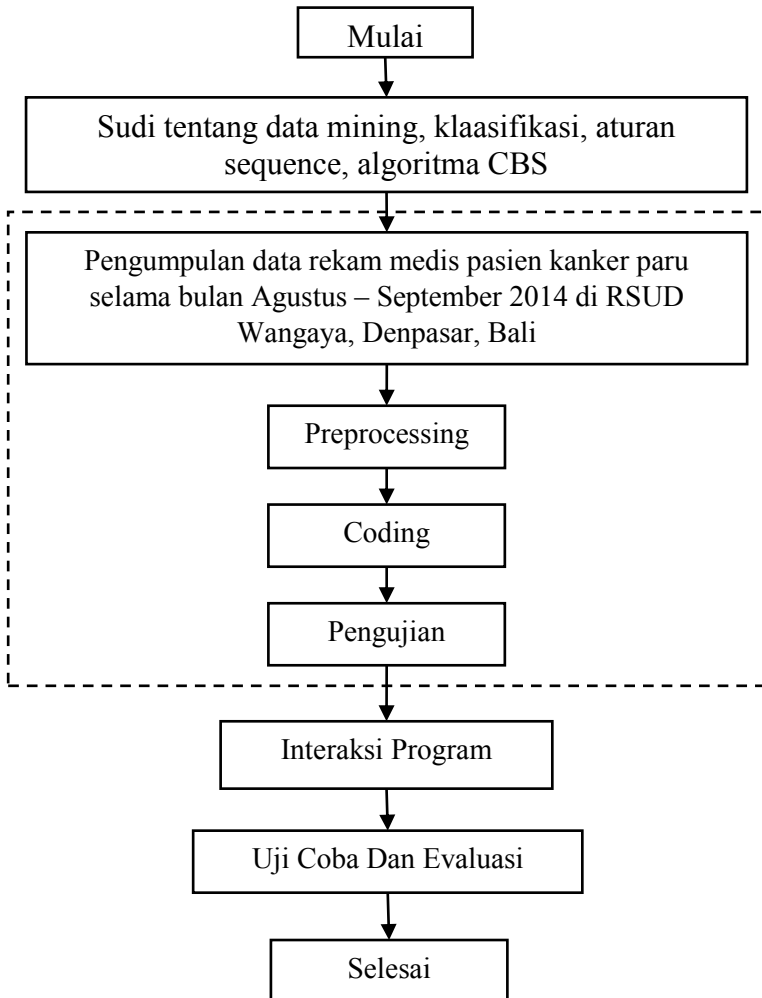
3.5 Penarikan Kesimpulan dan Saran

Pada tahap ini dilakukan penarikan kesimpulan terhadap hasil yang telah dicapai dari pengerjaan dan memberikan saran untuk pengembangan berikutnya.

3.6 Penyusunan Laporan Tugas Akhir

Merupakan tahap akhir dalam proses penelitian Tugas Akhir, dimana pada tahap ini akan ditulis semua hasil yang telah didapat selama mengerjakan tugas .

Berikut adalah diagram alir dari metodologi penelitian :



Gambar 3.1 Diagram Alir Penelitian

“Halaman ini sengaja dikosongkan”

BAB IV

ANALISA DAN PERANCANGAN PERANGKAT LUNAK

Pada bab ini akan dijelaskan mengenai analisa sistem dan perancangan perangkat lunak untuk Klasifikasi dengan aturan *Sequence* (CBS). Hal ini bertujuan untuk memberikan gambaran mengenai cara kerja sistem penghitungan akurasi pada algoritma CBS terhadap dataset kanker paru-paru dan bagaimana perancangan perangkat lunak yang akan dibuat. Analisa sistem meliputi proses CBS. Perancangan perangkat lunak meliputi perancangan proses dan perancangan antarmuka (*interface*).

4.1 Analisa Sistem

Analisa sistem memiliki tujuan memberi gambaran besar mengenai perangkat lunak yang hendak dibuat dan bagaimana interaksi dari sistem dengan pengguna. Dalam perangkat ini, program yang dibuat menggunakan aturan *Classify by Sequence* (CBS) dengan masukkan berupa data primer yang kemudian diolah menjadi data kategorial. Setelah data berbentuk data kategorial, selanjutnya akan dilakukan *preprocessing* dan memasukkan nilai *minimum support*. Keluaran dari program ini adalah nilai akurasi dari algoritma CBS dan besar cakupan data yang bisa diambil olgoritma CBS (dalam %).

Proses untuk mendapatkan Dataset pada perangkat lunak akan dijelaskan sebagai berikut:

1. Langkah awal pada program ini adalah memilih data asli yang akan dimasukkan ke dalam Matlab. Data asli tersebut adalah data mentah berupa file Microsoft Excel yang terlebih dahulu dikenai proses *data cleaning*, *data integration*, *data selection* dan *data transformation* sehingga data menjadi data kategorial. Lalu dilakukan *preprocessing* pada data tersebut. Langkah-langkah untuk mendapatkan dataset yang telah melalui *preprocessing* adalah sebagai berikut:

- a. Tekan tombol Browse untuk memilih dataset yang akan diolah, akan muncul beberapa dataset yang berbentuk file Microsoft Excel.

- b. Pilih salah satu dari dataset tersebut yang hendak *dipreprocessing*. Kemudian tekan Open.
 - c. Proses *preprocessing* secara otomatis dijalankan segera setelah pengguna membuka file yang akan diolah. Data mentah yang telah melalui proses *preprocessing* selanjutnya akan disimpan dalam sebuah matlab data.
2. Dilakukan *preprocessing* pada data mentah. Proses *preprocessing* dapat dilakukan secara otomatis dan manual. Pada program ini proses *preprocessing* dilakukan secara otomatis setelah data mentah dipilih. Pada proses ini program melakukan penyimpanan data ke dalam Matlab, menentukan kelas yang akan digunakan untuk proses CBS serta menentukan atribut-atribut yang nantinya digunakan untuk menentukan *rule* yang didapat dari *frequent item* (atribut yang sering muncul).
3. Setelah data selesai *dipreprocessing* langkah berikutnya adalah melakukan algoritma CBS pada data olahan.

4.2 Pengambilan dan Pengolahan Data

4.2.1 Pengambilan Data

Data yang digunakan merupakan data primer yang diperoleh dari Rumah Sakit Umum Daerah Wangaya, Denpasar, Bali selama bulan Agustus hingga akhir September 2014. Data yang digunakan adalah data rekam medis pasien kanker paru-paru dari bulan Januari sampai September tahun 2014 yang dapat dilihat pada Lampiran A.

Terdapat 15 atribut didalam data ini. Atribut-atribut yang ada di dalamnya antara lain:

1. Rontagen
2. Patologi Anatomis
3. Positif kanker paru
4. Batuk berdahak
5. Sesak nafas
6. Nyeri daerah dada
7. Demam

8. Nafsu makan menurun
9. Keringat malam
10. Suara serak
11. Lelah / lemas
12. Batuk berdarah
13. Berat badan turun
14. Benjolan pada pangkal leher
15. Pusing

4.2.2 Pengolahan Data

Beberapa faktor penyebab terjadinya kanker paru-paru antara lain:

1. Sesak nafas.
2. Batuk yang tidak kunjung sembuh.
3. Suara yang menciut saat menarik nafas.
4. Batuk berdarah.
5. Dahak yang dikeluarkan berwarna kuning pekat .
6. Penurunan berat badan secara drastic
7. Terlihat pembengkakan pada bagian leher ataupun wajah.

Sedemikian hingga atribut data yang digunakan untuk proses mencari *rule* dengan algoritma CBS pada Tugas Akhir ini meliputi positif kanker paru, batuk berdahak, sesak nafas, nyeri daerah dada, batuk berdarah dan patologi anatomis. Jumlah data yang digunakan untuk satu atribut adalah 80 buah *record* data.

Data mentah diatas kemudian dikenakan proses data *cleansing* untuk menghilangkan data yang tidak konsisten, baik nilai atribut maupun kelas. Lalu dilakukan data *integration*, yaitu menggabungkan beberapa sumber data menjadi satu. Selanjutnya dilakukan data *selection* dan data *transformation* sehingga hasil akhir dataset tersebut berupa data kategorial.

4.2.3 Preprocessing Data

Sebelum mencari *rule* dengan algoritma CBS dilakukan tahap *preprocessing* data terlebih dahulu. Disini tiap-tiap 80 buah *record* data dari masing-masing atribut tersebut akan disimpan.

4.3 Penemuan *Rule* Menggunakan Algoritma CBS

Pada Tugas Akhir ini algoritma CBS digunakan untuk menentukan *rule* pada data yang diolah di program ini seperti yang telah tertera pada bab 2. Tahapan Algoritma CBS yang dilakukan pertama adalah memisahkan data hasil *preprocessing* sejumlah 80% untuk *training* sehingga didapatkan *rule*. Untuk mendapatkan *rule* diperlukan adanya nilai *minimum support*, yaitu nilai sebagai nilai threshold untuk proses pemangkasan *frequent itemset*, dari dataset tersebut dicari nilai yang berbeda pada setiap nilai yang unik. Dari nilai tersebut dicari jumlahnya yang nilainya diambil dari dataset yang digunakan.

Karena pada tugas akhir ini akan dilihat nilai akurasi dari algoritma CBS untuk kanker paru, maka atribut kanker paru digunakan sebagai kelas.

Pada algoritma CBS untuk menemukan *rule* terdapat empat proses antara lain :

1. Menemukan CSP (*Classifiabel Sequential Pattern*).
2. Menemukan Frequent Itemset.
3. Membuat CSP Set.
4. Membuat Classifier.

Untuk prosedur algoritma diatas telah ditunjukkan pada Gambar 2.2.

4.3.1. Menemukan CSP.

Pada tahap ini semua CSP untuk klasifikasi temporal dari keseluruhan dataset. Dan dibandingkan antara kelas dan atribut, jika bernilai sama lalu disimpan pada variable sementara Dc1.

4.3.2. Menemukan *Frequent Itemset*.

Pada proses ini diciptakan variable untuk calon kandidat *frequent item* yang didapat dari nilai unik yang terdapat pada variable Dc1. Kemudian dicari jumlahnya. Dari hasil ini lalu dibandingkan dengan nilai *min_sup*.

4.3.3. Membuat CSP Set.

Setelah didapat nilai *Frequent Itemset*, CSP_Set didapat dengan mengumpulkan semua *Frequent Itemset*nya .

4.3.4. Membuat *Classifier*.

Setelah CSP_set didapatkan, semua *rule* yang tercipta dikumpulkan dan dibandingkan *rule* yang telah didapat dengan kelasnya dan diambil nilai *rule* yang memiliki nilai positif pada kelas dan atributnya..

4.4 Menentukan Nilai Akurasi.

Untuk menentukan nilai akurasi pada *rule* yang didapat, digunakan rumus :

Akurasi = (jum data yang benar / jum *rule* yang ada)*100%

Cakupan = (jum *rule* benar / jum data)*100%

4.5 Perancangan antarmuka.

Perancangan antar muka merupakan aspek penting yang perlu diperhatikan dalam pembuatan perangkat lunak. Pada umumnya antar muka harus bersifat user friendly sehingga mempermudah pengguna berinteraksi dengan system dan menjadi lebih nyaman dalam penngunaan perangkat lunaknya. Perancangan antarmuka pada perangkat lunak pada penelitian ini hanya terdiri dari satu form, yaitu:

1. Perancangan form awal klasifikasi kanker paru

Form ini merupakan form satu-satunya yang ditampilkan pada saat perangkat lunak ini dijalankan. Pada form ini berfungsi untuk memasukkan dataset yang akan digunakan untuk proses preprocessing dan juga untuk memasukan nilai *minimum support*. Lalu untuk hasil akurasi dan cakupan dari algoritma CBS terhadap keseluruhan dataset.

4.6 Implementasi antar muka.

Perangkat lunak yang dibangun pada Tugas Akhir ini terdiri dari satu form. Penjelasan dari form tersebut adalah sebagai berikut :

a. Tombol Browse

Tombol ini berfungsi untuk mencari dataset yang akan digunakan untuk masukkan.

b. TextBox File Path

TextBox ini berfungsi untuk menampilkan letak file yang telah dipilih sebagai masukkan

c. Tombol Start Process

Tombol ini berfungsi untuk memulai proses pencarian *rule* dengan menggunakan Algoritma CBS

d. TextBox Masukkan Nilai Minimum Support

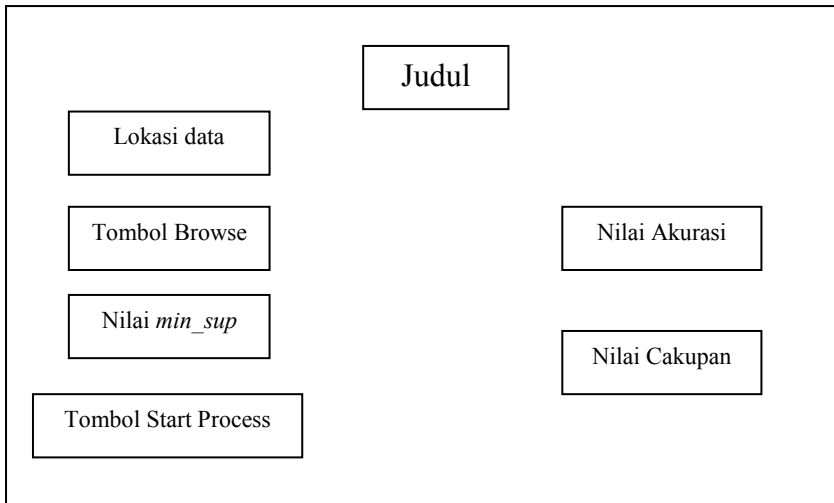
TextBox ini berfungsi untuk memasukkan nilai *min_sup* yang akan digunakan untuk proses CBS

e. TextBox Akurasi

TextBox ini berfungsi untuk menampilkan nilai akurasi yang didapat setelah proses CBS dilakukan.

f. TextBox Coverage

TextBox ini berfungsi untuk menampilkan nilai cakupan yang dihasilkan oleh *rule* yang dihasilkan oleh algoritma CBS.



Gambar 4.1 Perancangan Antarmuka

4.7 Perancangan kode program.

Langkah lanjutan yang dilakukan setelah proses analisa system, perancangan perangkat lunak dan desain antar muka adalah perancangan kode program. Perancangan ini terdiri atas beberapa proses antara lain :

1. Memasukkan nilai asli

Dalam melakukan klasifikasi berbasis aturan *sequence* diperlukan masukkan data yang berupa file Excel kedalam perangkat lunak untuk selanjutnya diproses. Untuk proses memasukkan data ditunjukkan pada Gambar 4.1 dan kode programnya ditampilkan pada Lampiran C

```
filename = uigetfile('*.xlsx');  
%Digunakan untuk mencegah error jk user tdk jd mengambil file  
if isequal(filename,0)  
    clear all;  
    clc  
    return  
end  
%Menampilkan nama file pada text1  
set(handles.edit1,'String',filename);
```

Gambar 4.2 Kode program memasukan data.

2. *Preprocessing Data*

Data yang telah dimasukkan pada perangkat lunak belum bisa langsung diproses, jadi diperlukan proses *preprocessing* data. Proses ini ditunjukkan pada Gambar 4.2A, 4.2B.

```
clear all;
clc;
time_start=tic;
a1=xlsread('book5.xlsx','A2:A81');
a2=xlsread('book5.xlsx','B2:B81');
a3=xlsread('book5.xlsx','C2:C81');
a4=xlsread('book5.xlsx','D2:D81');
a5=xlsread('book5.xlsx','E2:E81');
a6=xlsread('book5.xlsx','F2:F81');

m=length(a1);
for i1=1:m
    if a1(i1)==1
        class(i1)={'Kanker Paru'};
    else
        class(i1)={'Bukan Kanker Paru'};
    end
end

for i1=1:m
    if a2(i1)==1
        b1(i1)={'Batuk berdahak'};
    else
        b1(i1)={'Bukan Batuk berdahak'};
    end
end

for i1=1:m
    if a3(i1)==1
        b2(i1)={'Sesak'};
    else
        b2(i1)={'Bukan sesak'};
    end
end
```

Gambar 4.3A Kode Program Data *Preprocessing*.

```
for i1=1:m
    if a4(i1)==1
        b3(i1)={'Nyeri Daerah Dada'};
    else
        b3(i1)={'Bukan Nyeri Daerah Dada'};
    end
end

for i1=1:m
    if a5(i1)==1
        b4(i1)={'Batuk Berdarah'};
    else
        b4(i1)={'Bukan Batuk Berdarah'};
    end
end

for i1=1:m
    if a6(i1)==1
        b5(i1)={'Sitologi Klinis'};
    else
        b5(i1)={'Suptum'};
    end
end

atribut=[b1' b2' b3' b4' b5' ];
save atribut1 atribut;
save kelas class;
time end=toc(time start)
```

Gambar 4.3B Lanjutan Kode Program Data *Preprocessing*1.

3. Menemukan *frequent itemset*

Untuk menemukan *frequent itemset* ditunjukkan pada Gambar 4.3.

```
f_candidate(:,1)=unique(Dc1(:,1:5));
for j2=1:size(f_candidate,1)
    fre1(j2)=0;
    f1(j2)=0;
    for i4=1:N
        if sum(ismember(Dc1(i4,:),f_candidate(j2,1)))>0
            fre1(j2)=fre1(j2)+1;
        else
            fre1(j2)=fre1(j2);
        end
    end
    if fre1(j2)>=min_sup
        k=k+1;
        x=x+1;
        f1(k)=fre1(j2);
        F1(x)=f_candidate(j2);
        T1(k,1)=F1(x);
        CSP_SET(k,1)=F1(x);
        indeks_csp(k)=class(i1);
    end
end
```

Gambar 4.4 Kode Program Untuk Mencari Nilai *Frequent Itemset*.

4. Membangun *Classifier* menggunakan algoritma CBS.

Proses ini bertujuan untuk membangun *classifier* yang akan digunakan untuk melakukan perbandingan dengan data uji. Proses ini akan ditunjukkan pada Gambar 4.4.

```

for i3=1:size(CSP_SET,1)
    count(i3)=0;
    for i2=1:size(CSP_SET,1)
        if isequal(CSP_SET(i3,:),CSP_SET(i2,:)) == 1
            count(i3)=count(i3)+1;
        end
    end
    if count(i3) == size(class,1)
        k=k+1;
        csp_dup(k,:) = CSP_SET(i3,:);
    end
end
k=k-k;
a(1,1:size(CSP_SET,1)) = {' '};
h=exist('csp_dup');
if h==1
    for i5=1:size(csp_dup)
        for j3=1:size(CSP_SET)
            if isequal(csp_dup(i5,:),CSP_SET(j3,:)) == 1
                CSP_SET(j3,:)=a;
            end
        end
    end
end
for j3=1:size(CSP_SET,1)
    if isequal(CSP_SET(j3,:),a) == 0
        k=k+1;
        T(k,1:size(CSP_SET,1)) = CSP_SET(j3,:);
        T(k,size(CSP_SET,1)+1) = {'>'};
        T(k,size(CSP_SET,1)+2) = indeks_csp(j3);
    end
end
end

```

Gambar 4.5 Kode Program Untuk Membangun *Classifier*.

5. Mencari nilai akurasi dan coverage.

Proses ini bertujuan untuk mencari nilai akurasi dari *rule* yang ditemukan dengan pembandingan data uji yang telah ditentukan.

```

clear all;
clc;
%count accuracy
load kelas;
load atribut1;
load cspmodif;
jum_train=0.2*length(class);
D=[atribut class'];
for p=1:jum_train
    Dtes(p,:)=D(p,:);
    Dbodi(p,:)=D(p,:);
end
x=0;countbodi=0;
counttes=0;
for k=1:size(T,1)
    for kk=1:size(T',1)-3
        if isequal(T(k,kk),{[]})==0
            t=T(k,1:kk);
            for jj=1:jum_train
                if sum(ismember(Dtes(jj,:),t))>0&&isequal(T(k,size(T',1)),Dtes(jj,size(Dtes',1)))==1
                    counttes=counttes+1;
                    Dtes(jj,:)={' '};
                end
                if sum(ismember(Dbodi(jj,:),t))>0
                    countbodi=countbodi+1;
                    Dbodi(jj,:)={' '};
                end
            end
        end
    end
end
t={};
end
coverage=(countbodi/jum_train)*100
accuracy=(counttes/countbodi)*100
covacc=(accuracy/100)*(1+(coverage/100))

```

Gambar 4.6 Kode Program Untuk Mencari Akurasi

“ Halaman ini sengaja dikosongkan “

BAB V

PEMBAHASAN DAN PENGUJIAN

Pada bab ini akan dilakukan uji coba program klasifikasi berdasarkan aturan *sequence* menggunakan algoritma CBS dengan beberapa jenis nilai *minimum support* yang berbeda. Pengujian dilakukan untuk mengetahui apakah program ini menghasilkan nilai keluaran sesuai dengan tujuan yang telah disebutkan di awal

5.1 Uji coba

Pengujian dilakukan menggunakan 20% dari total data keseluruhan, yaitu sekitar 16 data yang digunakan. Hasil yang didapat pada proses ini adalah rule dan nilai akurasi serta coverage(cakupan) dari algoritma CBS terhadap dataset. Tampilan untuk nilai `min_sup` 0.9 beserta rule yang diperoleh akan ditampilkan pada Gambar 5.2 dan 5.3, karena untuk proses preprocessing dilakukan secara otomatis maka setelah dataset dimasukkan pengguna akan diminta untuk memasukkan nilai `minimum support (min_sup)` yang dikehendaki.



Gambar 5.1 Hasil uji coba untuk tampilan awal



Gambar 5.2 Hasil uji coba untuk nilai min_sup 0.9

	1	2	3
1	Suptum	=>	Bukan Kanker Paru
2	Sitologi Klinis	=>	Kanker Paru
3			

Gambar 5.3 Rule pada saat nilai min_sup 0.9

Pada nilai min_sup 0.9, nilai akurasi masih belum didapatkan, dikarenakan rule yang didapat belum dapat mencakup keseluruhan data. Sehingga algoritma CBS tidak dapat membuat classifier yang sesuai dengan pola pada dataset. Selanjutnya nilai minimum support akan diubah menjadi 0.8 untuk melihat apakah algoritma CBS dapat membuat rule yang mencakup keseluruhan dataset.

Untuk hasil running program akan ditampilkan pada Gambar 5.4 dan rule yang dihasilkan akan ditampilkan pada Gambar 5.5.



Gambar 5.4 Hasil uji coba untuk nilai min_sup 0.8

T <2x3 cell>			
	1	2	3
1	Suptum	=>	Bukan Kanker Paru
2	Sitologi Klinis	=>	Kanker Paru
3			

Gambar 5.5 Rule pada saat nilai min_sup 0.8

Pada nilai min_sup 0.8 juga menghasilkan nilai yang sama seperti yang dihasilkan pada nilai min_sup 0.9. program masih belum dapat menentukan rule yang dapat mencakup keseluruhan

dataset. Maka nilai dari minimum support diturunkan kembali menjadi 0.7. untuk hasil running pada min_sup ini akan ditunjukkan pada Gambar 5.6 dan rule yang dihasilkan ditunjukkan pada Gambar 5.7.

KLASIFIKASI KANKER PARU DENGAN MENGGUNAKAN ALGORITMA CLASSIFY BY SEQUENCE

Data Set

Minimum Support

Accuracy Level (%)

Coverage (%)

I Putu Eddy Sukadiana
1206100711

Gambar 5.6 Hasil uji coba untuk nilai min_sup 0.7

T <4x4 cell>				
	1	2	3	4
1	'Suptum'	[]	'=>'	'Bukan Kan...
2	'Bukan Batuk Berdarah'	[]	'=>'	'Kanker Paru'
3	'Sitologi Klinis'	[]	'=>'	'Kanker Paru'
4	'Bukan Batuk Berdarah'	'Sitologi Klinis'	'=>'	'Kanker Paru'
5				

Gambar 5.7 Rule pada saat nilai min_sup 0.7

Pada nilai min_sup 0.7 didapat nilai akurasi 100% dan nilai coverage hingga 100% ini berarti rule yang diperoleh pada minimum support ini dapat mencakup keseluruhan dataset. Jumlah rule yang diperoleh juga semakin banyak, sehingga memungkinkan algoritma CBS untuk mengklasifikasikan seluruh dataset. Secara kedokteran keputusan akhir untuk menentukan ada tidaknya sel kanker pada sediaan yang dikirim ke laboratorium, namun pada rule yang didapat pada nilai min_sup 0.7 diatas terdapat keanehan yaitu pada atribut bukan batuk berdarah langsung diklasifikasikan sebagai kanker, padahal pada analisa kedokteran hal ini belum dapat memutuskan seseorang menderita kanker paru. Jadi nilai minimum support diturunkan menjadi 0.4 untuk melihat rule yang dihasilkan, untuk running pada nilai min_sup 0.4 ditunjukkan pada Gambar 5.8 dan rule yang didapat ditunjukkan pada Gambar 5.9A dan Gambar 5.9B.



Gambar 5.8 Hasil uji coba untuk nilai min_sup 0.4

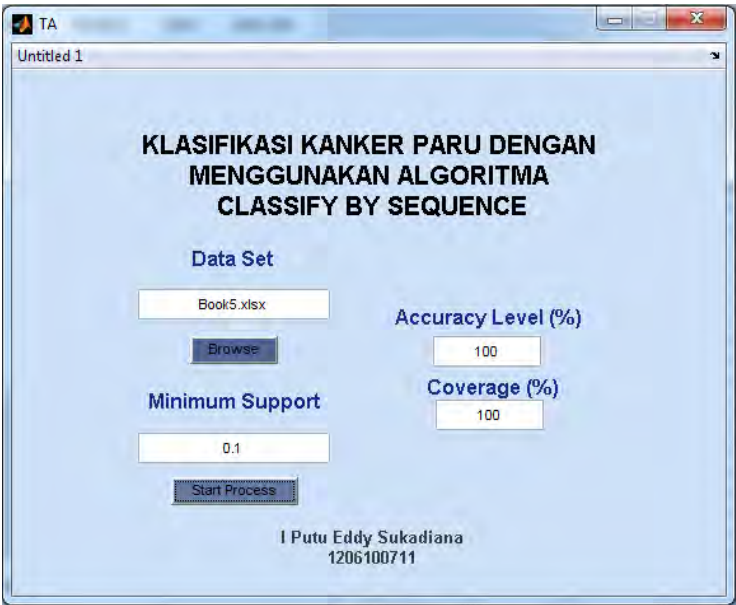
T <32x6 cell>						
	1	2	3	4	5	6
1	'Batuk Berdarah'	()	()	()	'=>'	'Bukan Kanker Paru'
2	'Bukan sesak'	()	()	()	'=>'	'Bukan Kanker Paru'
3	'Suptum'	()	()	()	'=>'	'Bukan Kanker Paru'
4	'Batuk Berdarah'	'Suptum'	()	()	'=>'	'Bukan Kanker Paru'
5	'Bukan Batuk Berdarah'	'Suptum'	()	()	'=>'	'Bukan Kanker Paru'
6	'Bukan Batuk berdahak'	'Batuk Berdarah'	()	()	'=>'	'Bukan Kanker Paru'
7	'Bukan Batuk berdahak'	'Bukan Nyeni Daerah Dada'	()	()	'=>'	'Bukan Kanker Paru'
8	'Bukan Batuk berdahak'	'Suptum'	()	()	'=>'	'Bukan Kanker Paru'
9	'Bukan Nyeni Daerah Dada'	'Suptum'	()	()	'=>'	'Bukan Kanker Paru'
10	'Bukan sesak'	'Suptum'	()	()	'=>'	'Bukan Kanker Paru'
11	'Nyeni Daerah Dada'	'Suptum'	()	()	'=>'	'Bukan Kanker Paru'
12	'Bukan Batuk berdahak'	'Batuk Berdarah'	'Suptum'	()	'=>'	'Bukan Kanker Paru'
13	'Bukan Batuk berdahak'	'Bukan Nyeni Daerah Dada'	'Suptum'	()	'=>'	'Bukan Kanker Paru'

Gambar 5.9A Rule pada saat nilai min_sup 0.4

14	'Batuk berdahak'	()	()	()	'=>'	'Kanker Paru'
15	'Sesak'	()	()	()	'=>'	'Kanker Paru'
16	'Sitologi Klinis'	()	()	()	'=>'	'Kanker Paru'
17	'Batuk berdahak'	'Bukan Batuk Berdarah'	()	()	'=>'	'Kanker Paru'
18	'Batuk berdahak'	'Sesak'	()	()	'=>'	'Kanker Paru'
19	'Batuk berdahak'	'Sitologi Klinis'	()	()	'=>'	'Kanker Paru'
20	'Bukan Batuk Berdarah'	'Sitologi Klinis'	()	()	'=>'	'Kanker Paru'
21	'Bukan Batuk berdahak'	'Sitologi Klinis'	()	()	'=>'	'Kanker Paru'
22	'Bukan Nyeri Daerah Dada'	'Sitologi Klinis'	()	()	'=>'	'Kanker Paru'
23	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	()	()	'=>'	'Kanker Paru'
24	'Nyeri Daerah Dada'	'Sitologi Klinis'	()	()	'=>'	'Kanker Paru'
25	'Sesak'	'Bukan Batuk Berdarah'	()	()	'=>'	'Kanker Paru'
26	'Sesak'	'Sitologi Klinis'	()	()	'=>'	'Kanker Paru'
27	'Batuk berdahak'	'Bukan Batuk Berdarah'	'Sitologi Klinis'	()	'=>'	'Kanker Paru'
28	'Batuk berdahak'	'Sesak'	'Bukan Batuk Berdarah'	()	'=>'	'Kanker Paru'
29	'Batuk berdahak'	'Sesak'	'Sitologi Klinis'	()	'=>'	'Kanker Paru'
30	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Sitologi Klinis'	()	'=>'	'Kanker Paru'
31	'Sesak'	'Bukan Batuk Berdarah'	'Sitologi Klinis'	()	'=>'	'Kanker Paru'
32	'Batuk berdahak'	'Sesak'	'Bukan Batuk Berdarah'	'Sitologi Klinis'	'=>'	'Kanker Paru'

Gambar 5.9B Rule pada saat nilai min_sup 0.4

Pada nilai minimum support 0.4 menunjukkan bahwa rule yang dihasilkan semakin banyak dan nilai akurasi tetap 100% namun jika rule yang dihasilkan dilihat dengan teliti maka terdapat pula beberapa hasil yang bisa dikatakan masih belum cukup untuk menentukan seseorang menderita kanker paru ataupun tidak. Maka dilakukan uji coba lanjutan untuk melihat apakah hasil rule masih menunjukkan nilai yang serupa seperti pada rule yang dihasilkan pada nilai min_sup sebelumnya. Untuk uji coba berikutnya digunakan nilai min_sup 0.1. untuk hasil running program ditunjukkan pada Gambar 5.10 dan rule yang diperoleh ditunjukkan pada Gambar 5.11A, 5.11B, 5.11C, dan 5.11D.



Gambar 5.10 Hasil uji coba untuk nilai min_sup 0.1

T <120>? .gali>							
	1	2	3	4	5	6	7
1	Suptom					es>	Bukan Kan...
2	Batuk Berdahak	Suptom				es>	Bukan Kan...
3	Batuk berdahak	Suptom				es>	Bukan Kan...
4	Bukan Batuk Berdahak	Suptom				es>	Bukan Kan...
5	Bukan Batuk berdahak	Suptom				es>	Bukan Kan...
6	Bukan Nyeni Daerah Dada	Suptom				es>	Bukan Kan...
7	Bukan sesak	Suptom				es>	Bukan Kan...
8	Nyeni Daerah Dada	Batuk Berdahak				es>	Bukan Kan...
9	Nyeni Daerah Dada	Suptom				es>	Bukan Kan...
10	Sesak	Batuk Berdahak				es>	Bukan Kan...
11	Sesak	Suptom				es>	Bukan Kan...
12	Batuk berdahak	Bukan Batuk Berdahak	Suptom			es>	Bukan Kan...
13	Batuk berdahak	Bukan Nyeni Daerah Dada	Suptom			es>	Bukan Kan...
14	Batuk berdahak	Bukan sesak	Bukan Nyeni Daerah Dada			es>	Bukan Kan...
15	Batuk berdahak	Bukan sesak	Suptom			es>	Bukan Kan...
16	Batuk berdahak	Nyeni Daerah Dada	Suptom			es>	Bukan Kan...
17	Bukan Batuk berdahak	Batuk Berdahak	Suptom			es>	Bukan Kan...
18	Bukan Batuk berdahak	Bukan Batuk Berdahak	Suptom			es>	Bukan Kan...
19	Bukan Batuk berdahak	Bukan Nyeni Daerah Dada	Suptom			es>	Bukan Kan...
20	Bukan Batuk berdahak	Bukan sesak	Bukan Nyeni Daerah Dada			es>	Bukan Kan...
21	Bukan Batuk berdahak	Bukan sesak	Suptom			es>	Bukan Kan...
22	Bukan Batuk berdahak	Nyeni Daerah Dada	Batuk Berdahak			es>	Bukan Kan...
23	Bukan Batuk berdahak	Nyeni Daerah Dada	Suptom			es>	Bukan Kan...
24	Bukan Batuk berdahak	Sesak	Batuk Berdahak			es>	Bukan Kan...
25	Bukan Batuk berdahak	Sesak	Bukan Nyeni Daerah Dada			es>	Bukan Kan...
26	Bukan Batuk berdahak	Sesak	Suptom			es>	Bukan Kan...
27	Bukan Nyeni Daerah Dada	Batuk Berdahak	Suptom			es>	Bukan Kan...
28	Bukan Nyeni Daerah Dada	Bukan Batuk Berdahak	Suptom			es>	Bukan Kan...
29	Bukan sesak	Batuk Berdahak	Suptom			es>	Bukan Kan...
30	Bukan sesak	Bukan Batuk Berdahak	Suptom			es>	Bukan Kan...
31	Bukan sesak	Bukan Nyeni Daerah Dada	Batuk Berdahak			es>	Bukan Kan...

Gambar 5.11A Rule pada saat nilai min_sup 0.1

T <1267 call>

1	2	3	4	5	6	7
31. 'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'				'Bukan Kan...
32. 'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'				'Bukan Kan...
33. 'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Supturn'				'Bukan Kan...
34. 'Bukan sesak'	'Nyeri Daerah Dada'	'Batuk Berdarah'				'Bukan Kan...
35. 'Bukan sesak'	'Nyeri Daerah Dada'	'Supturn'				'Bukan Kan...
36. 'Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'				'Bukan Kan...
37. 'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Supturn'				'Bukan Kan...
38. 'Sesak'	'Batuk Berdarah'	'Supturn'				'Bukan Kan...
39. 'Sesak'	'Bukan Batuk Berdarah'	'Supturn'				'Bukan Kan...
40. 'Sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'				'Bukan Kan...
41. 'Sesak'	'Bukan Nyeri Daerah Dada'	'Supturn'				'Bukan Kan...
42. 'Sesak'	'Nyeri Daerah Dada'	'Supturn'				'Bukan Kan...
43. 'Batuk berdarah'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Supturn'			'Bukan Kan...
44. 'Batuk berdarah'	'Bukan sesak'	'Bukan Batuk Berdarah'	'Supturn'			'Bukan Kan...
45. 'Batuk berdarah'	'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'			'Bukan Kan...
46. 'Batuk berdarah'	'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Supturn'			'Bukan Kan...
47. 'Batuk berdarah'	'Bukan sesak'	'Nyeri Daerah Dada'	'Supturn'			'Bukan Kan...
48. 'Bukan Batuk berdarah'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'			'Bukan Kan...
49. 'Bukan Batuk berdarah'	'Bukan sesak'	'Batuk Berdarah'	'Supturn'			'Bukan Kan...
50. 'Bukan Batuk berdarah'	'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'			'Bukan Kan...
51. 'Bukan Batuk berdarah'	'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Supturn'			'Bukan Kan...
52. 'Bukan Batuk berdarah'	'Bukan sesak'	'Nyeri Daerah Dada'	'Supturn'			'Bukan Kan...
53. 'Bukan Batuk berdarah'	'Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'			'Bukan Kan...
54. 'Bukan Batuk berdarah'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Supturn'			'Bukan Kan...
55. 'Bukan Batuk berdarah'	'Sesak'	'Batuk Berdarah'	'Supturn'			'Bukan Kan...
56. 'Bukan Batuk berdarah'	'Sesak'	'Bukan Batuk Berdarah'	'Supturn'			'Bukan Kan...
57. 'Bukan Batuk berdarah'	'Sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'			'Bukan Kan...
58. 'Bukan Batuk berdarah'	'Sesak'	'Bukan Nyeri Daerah Dada'	'Supturn'			'Bukan Kan...
59. 'Bukan Batuk berdarah'	'Sesak'	'Nyeri Daerah Dada'	'Supturn'			'Bukan Kan...
60. 'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'			'Bukan Kan...
61. 'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Supturn'			'Bukan Kan...

Gambar 5.11B Rule pada saat nilai min_sup 0.1

1	2	3	4	5	6	7
61. 'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Supturn'			'Bukan Kan...
62. 'Bukan sesak'	'Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'			'Bukan Kan...
63. 'Bukan sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Supturn'			'Bukan Kan...
64. 'Sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'			'Bukan Kan...
65. 'Batuk berdarah'	'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Supturn'		'Bukan Kan...
66. 'Bukan Batuk berdarah'	'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'		'Bukan Kan...
67. 'Bukan Batuk berdarah'	'Sesak'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'	'Supturn'		'Bukan Kan...
68. 'Sitologi Klinis'						'Kanker Panu'
69. 'Batuk Berdarah'	'Sitologi Klinis'					'Kanker Panu'
70. 'Batuk berdarah'	'Sesak'					'Kanker Panu'
71. 'Batuk berdarah'	'Sitologi Klinis'					'Kanker Panu'
72. 'Bukan Batuk Berdarah'	'Sitologi Klinis'					'Kanker Panu'
73. 'Bukan Batuk berdarah'	'Sitologi Klinis'					'Kanker Panu'
74. 'Bukan Nyeri Daerah Dada'	'Sitologi Klinis'					'Kanker Panu'
75. 'Bukan sesak'	'Sitologi Klinis'					'Kanker Panu'
76. 'Nyeri Daerah Dada'	'Sitologi Klinis'					'Kanker Panu'
77. 'Sesak'	'Sitologi Klinis'					'Kanker Panu'
78. 'Batuk berdarah'	'Bukan Batuk Berdarah'	'Sitologi Klinis'				'Kanker Panu'
79. 'Batuk berdarah'	'Bukan Nyeri Daerah Dada'	'Sitologi Klinis'				'Kanker Panu'
80. 'Batuk berdarah'	'Bukan sesak'	'Sitologi Klinis'				'Kanker Panu'
81. 'Batuk berdarah'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'				'Kanker Panu'
82. 'Batuk berdarah'	'Nyeri Daerah Dada'	'Sitologi Klinis'				'Kanker Panu'
83. 'Batuk berdarah'	'Sesak'	'Bukan Batuk Berdarah'				'Kanker Panu'
84. 'Batuk berdarah'	'Sesak'	'Bukan Nyeri Daerah Dada'				'Kanker Panu'
85. 'Batuk berdarah'	'Sesak'	'Nyeri Daerah Dada'				'Kanker Panu'
86. 'Batuk berdarah'	'Sesak'	'Sitologi Klinis'				'Kanker Panu'
87. 'Bukan Batuk berdarah'	'Batuk Berdarah'	'Sitologi Klinis'				'Kanker Panu'
88. 'Bukan Batuk berdarah'	'Bukan Batuk Berdarah'	'Sitologi Klinis'				'Kanker Panu'
89. 'Bukan Batuk berdarah'	'Bukan Nyeri Daerah Dada'	'Sitologi Klinis'				'Kanker Panu'
90. 'Bukan Batuk berdarah'	'Bukan sesak'	'Sitologi Klinis'				'Kanker Panu'
91. 'Bukan Batuk berdarah'	'Nyeri Daerah Dada'	'Sitologi Klinis'				'Kanker Panu'

Gambar 5.11C Rule pada saat nilai min_sup 0.1

	1	2	3	4	5	6	7
92	'Bukan Batuk berdarah'	'Sesak'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
93	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
94	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
95	'Bukan sesak'	'Batuk Berdarah'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
96	'Bukan sesak'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
97	'Bukan sesak'	'Bukan Nyeri Daerah Dada'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
98	'Bukan sesak'	'Nyeri Daerah Dada'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
99	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
100	'Sesak'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
101	'Sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	[]	[]	'=s'	'Kanker Paru'
102	'Sesak'	'Bukan Nyeri Daerah Dada'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
103	'Sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	[]	[]	'=s'	'Kanker Paru'
104	'Sesak'	'Nyeri Daerah Dada'	'Stetologi Klinis'	[]	[]	'=s'	'Kanker Paru'
105	'Batuk berdarah'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
106	'Batuk berdarah'	'Bukan sesak'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
107	'Batuk berdarah'	'Bukan sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	[]	'=s'	'Kanker Paru'
108	'Batuk berdarah'	'Bukan sesak'	'Nyeri Daerah Dada'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
109	'Batuk berdarah'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
110	'Batuk berdarah'	'Sesak'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
111	'Batuk berdarah'	'Sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	[]	'=s'	'Kanker Paru'
112	'Batuk berdarah'	'Sesak'	'Bukan Nyeri Daerah Dada'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
113	'Batuk berdarah'	'Sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	[]	'=s'	'Kanker Paru'
114	'Batuk berdarah'	'Sesak'	'Nyeri Daerah Dada'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
115	'Bukan Batuk berdarah'	'Bukan Nyeri Daerah Dada'	'Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
116	'Bukan Batuk berdarah'	'Bukan sesak'	'Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
117	'Bukan Batuk berdarah'	'Bukan sesak'	'Nyeri Daerah Dada'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
118	'Bukan Batuk berdarah'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
119	'Bukan Batuk berdarah'	'Sesak'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
120	'Bukan Batuk berdarah'	'Sesak'	'Nyeri Daerah Dada'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
121	'Bukan sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
122	'Sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
123	'Sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	[]	'=s'	'Kanker Paru'
124	'Batuk berdarah'	'Bukan sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	'=s'	'Kanker Paru'
125	'Batuk berdarah'	'Sesak'	'Bukan Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	'=s'	'Kanker Paru'
126	'Batuk berdarah'	'Sesak'	'Nyeri Daerah Dada'	'Bukan Batuk Berdarah'	'Stetologi Klinis'	'=s'	'Kanker Paru'
127							

Gambar 5.11D Rule pada saat nilai min_sup 0.1

Dari hasil percobaan pada nilai min_sup 0.1 didapat nilai akurasi yang juga mencapai 100% namun beberapa rule yang masih belum dapat memutuskan seseorang menderita kanker paru tetap muncul meskipun algoritma CBS telah menghasilkan sebanyak 126 rule.

Hasil running program untuk setiap nilai min_sup dari 0.1 hingga 0.9 akan ditampilkan pada tabel dibawah ini :

No	<i>Min_sup</i>	Accuracy	Coverage	Waktu
1	0.9	-	-	4.01
2	0.8	-	0	4.47
3	0.7	100	100	2.79
4	0.6	100	100	4.70
5	0.5	100	100	8.35
6	0.4	100	100	10.70
7	0.3	100	100	18.85
8	0.2	100	100	49.77
9	0.1	100	100	122.99

Table 5.1 Nilai akurasi, cakupan dan waktu proses tiap *min_sup*.

BAB VI PENUTUP

6.1 Kesimpulan

Berdasarkan hasil dari penelitian tentang klasifikasi kanker paru dengan menggunakan algoritma CBS ini, maka didapatkan kesimpulan berikut :

1. Klasifikasi terhadap kanker paru dapat dilakukan dengan menggunakan algoritma CBS, sehingga pasien dapat diklasifikasikan dengan mudah.
2. Algoritma CBS dapat diterapkan pada dataset kecil, namun nilai akurasi yang didapat masih belum dapat diandalkan.
3. Semakin kecil nilai minimum support, maka nilai akurasi algoritma CBS semakin menurun.
4. Semakin kecil nilai minimum support maka waktu yang digunakan untuk mengolah data akan semakin lama, karena semakin banyaknya rule yang diperoleh.

6.2 Saran

Dengan melihat hasil yang dicapai pada penelitian ini, ada beberapa hal yang penulis sarankan untuk pengembangan selanjutnya, yaitu :

1. Metode ini dapat dikembangkan tidak sebatas hanya untuk rekam medik pasien kanker paru saja tetapi juga pada rekam medis penyakit lainnya yang bersangkutan dengan nilai patologi klinis dan patologi anatomisnya.
2. Perlu adanya peningkatan pada efisiensi algoritma CBS beserta akurasinya terhadap dataset kecil.

"Halaman ini sengaja dikosongkan"

DAFTAR PUSTAKA

- [1] Han, J., dkk. (2005). *“Sequential Pattern Mining By Pattern-Growth : Principles And Extension”*. StudFuzz.
- [2] Hsu, W. Dkk. (2008). *“Temporal And Spatio-Temporal Data mining”*, IGI Global., Hershey dan London.
- [3] Mastrogianis, N. dkk. (2009). *“A Method For Improving The Accuracy of Data mining Classification Algorithm”*. *Journal of Computer And Operation Research*.
- [4] Mitsa, C., (2010), *“Temporal Data mining”*, A Chapman & Hall/CRC., New York.
- [5] Tan, N.P. dkk, (2006), *“Introduction to Data mining”*, Pearson AddisonWeasly., New york.
- [6] Tseng, V. S. dan Lee, C. H. (2009). *“Effective Temporal Data Classification By Integrating Sequential Pattern Mining And Probabilistic Induction”*, *Journal of Expert System with Application*.
- [7] Han, J., dkk. (2006). *“Data mining: Concepts and Techniques”*, Second edition,
- [8] Iqbal, Muhammad. (2012). *“ Peningkatan Efisiensi Pruning Pada Algoritma CBS Menggunakan Algoritma FEAT”*. Thesis Magister Matematika, Institut Teknologi Sepuluh Nopember.
- [9] Santosa, Budi. (2007). *“ Data mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori & Aplikasi “*. Graha Ilmu, Yogyakarta.
- [10] Tan, N.P. Steinbach, M., Kumar, V, (2006), *“Introduction to Data mining”*, Pearson AddisonWeasly., New York.
- [11] <http://yayasankankerindonesia.org/tentang-kanker/>.
- [12] Ahmed, Kawasar, (2013), *“Early Detection of Lung Cancer Risk Using Data mining”*. *Asian Pasific j Cancer Prev*, page 595-598.

- [13] Ada, dkk (2013). “ *A Study of Detection of Lung Cancer Using Data mining Classification Techniques*”. *International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 3* page 131-134.
- [14] Sharma, P dan Ojha, P (2012), “ *Classification By Clustering Based On Adjusted Cluster*”, *Journal of Computer Engineering*.
- [15] www.kompas.com/health/read/2012/02/11/07151724/Kanker.Paru.Naik.

LAMPIRAN

A. Cara Menjalankan Perangkat Lunak

Untuk dapat menjalankan perangkat lunak ini langkah-langkah yang harus dilakukan adalah sebagai berikut.

1. Jalankan program Matlab dan jalankan file TA.fig.
2. Tekan *push button* Browse dan pilih dataset yang akan diproses.
3. Masukkan nilai *minimum support* dengan rentang 0,1-0,9. Dimana nilai minimum support dibatasi satu angka dibelakang koma.
4. Mulai proses *preprocessing*. Maka nilai cakupan dan akurasi akan diperoleh setelah proses selesai..

B. Dataset yang Digunakan

Berikut adalah table dari data yang didapat dari RSUD Wangaya, Denpasar , Bali.

Tabel B. Dataset

No	No RM	Patologi Klinis
1	420388	batuk, sesak, nafsu makan turun
2	427480	nyeri dada, lelah, batuk
3	192210	demam, batuk, nyeri dada
4	461414	nyeri dada, sesak
5	917539	
6	924565	
7	931592	
8	556626	nyeri dada kanan, cepat lelah
9	896459	
10	903486	
11	910512	
12	422174	batuk, sesak
13	462626	batuk berdarah

No	No RM	Patologi Klinis
14	462876	batuk, sesak, nyeri di dada
15	847274	
16	854300	
17	861326	
18	868353	
19	875380	
20	882406	
21	889433	
22	525798	sesak, batuk berdahak, demam
23	523665	lemas, sesak, batuk
24	341056	sesak, batuk berdahak, demam
25	503448	batuk, sesak, nyeri ulu hati
26	464918	batuk berdarah, keringat malam
27	525155	batuk, nyeri dada, sesak,
28	480660	batuk, sesak, nyeri dada
29	473235	sesak, batuk, demam
30	479775	sesak, batuk
31	784035	
32	791062	
33	798088	
34	805115	
35	812141	
36	819168	
37	826194	
38	833221	
39	840247	
40	481546	batuk darah, sesak
41	407657	batuk darah, pusing, sesak

No	No RM	Patologi Klinis
42	498315	sesak, batuk berdarah
43	506663	sesak, batuk, nafsu makan turun
44	501778	nyeri ulu hati, sesak
45	47498	batuk, demam, nyeri dada
46	480491	sesak, batuk, nyeri dada
47	748903	nafsu makan turun, batuk lama, demam
48	755929	batuk berdarah, demam
49	762956	
50	769982	
51	777009	
52	538991	nyeri ulu hati, batuk berdarah
53	576901	batuk, nyeri uluhati, berat badan turun
54	493471	nyeri dada, sesak
55	558410	batuk berdarah, keringat malam
56	473845	batuk, sesak, keringat malam
57	872025	nyeri dada, sesak, batuk berdarah
58	480313	batuk, nyeri uluhati
59	713903	nyeri dada, batuk
60	784671	sesak, keringat malam
61	760216	nyeri dada, demam, sesak
62	446124	batuk berdahak, sesak, demam
63	773929	batuk berdarah, keringat malam, demam
64	719422	nyeri ulu hati, batuk berdarah, demam
65	814313	batuk berdarah, demam, benjolan pangkal leher

No	No RM	Patologi Klinis
66	727823	batuk berdahak,demam, keringat malam
67	734850	keringat malam, nafsu makan turun, batuk lama
68	741876	nyeri ulu hati, batuk berdarah
69	490371	benjolan pangkal leher, sulit menelan, batuk
70	846028	batuk, sembab muka, nafsu makan menurun
71	475914	batuk berdarah
72	410993	nyeri dada, sesak
73	831922	benjolan pangkal leher, sesak, batuk
74	882616	nyeri dada, suara serak, benjol pangkal leher
75	819210	batuk, demam, sesak
76	831990	sesak, batuk, susah menelan, keringat malam, demam
77	847513	sakit dada, suara serak, benjolan pangkal leher
78	459750	batuk, nyeri uluhati, sulit menelan
79	534290	batuk, sesak, nafsu makan turun
80	347801	demam, batuk, nyeri dada

Lanjutan tabel

Rontagen	Patologi Anatomis
efusi pleura massif kanan	sitologi : NSCC
efusi pleura massif kanan	sitologi : NSCC
efusi pleura massif kanan	sitologi : SCC
efusi pleura kanan	sitologi : SCC
efusi pleura unilateral, infiltrat	sputum : BTA ++
efusi pleura unilateral, infiltrat, kaviti dengan dinding tebal	sputum : BTA +++
efusi pleura unilateral, infiltrat	sputum : BTA ++, segmen posterior lobus superior
efusi pleura kanan	sitologi : SCC
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral, infiltrat	sputum : BTA ++, segmen posterior lobus superior
efusi pleura unilateral, infiltrat	sputum : BTA +++
efusi pleura massif kiri	sitologi : NSCC
efusi pleura bilateral	sitologi : NSCC
efusi pleura	metostatik non keratinizing squamous cell carcinoma
efusi pleura unilateral	sputum : BTA ++, segmen posterior lobus superior
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral	sputum : BTA ++
efusi pleura unilateral, infiltrat, kaviti dengan dinding tebal	sputum : BTA ++
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral, infiltrat, kaviti dengan dinding tebal	sputum : BTA ++

Rontagen	Patologi Anatomis
efusi pleura unilateral	sputum : BTA ++, segmen posterior lobus superior
efusi pleura	sitologi : SCC
efusi pleura bilateral massif	sitologi : NSCC
efusi pleura	sitologi : SCC
efusi pleura massif kiri	sitologi : NSCC
efusi pleura	sitologi : SCC
efusi pleura	sitologi : SCC
efusi pleura kanan	sitologi : SCC
efusi pleura kanan	sitologi : SCC
efusi pleura kiri	sitologi : SCC
efusi pleura unilateral, infiltrat, kaviti dengan dinding tebal	sputum : BTA +++
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral, infiltrat	sputum : BTA ++
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral, infiltrat, kaviti dengan dinding tebal	sputum : BTA ++, segmen posterior lobus superior
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral, infiltrat	sputum : BTA ++
efusi pleura massif kanan	sitologi : NSCC
efusi pleura bilateral	sitologi : NSCC
efusi pleura, pulmo paru kiri	sitologi : NSCC
efusi bilateral	sitologi : NSCC
efusi pleura massif kanan	sitologi : NSCC
efusi pleura	sitologi : NSCC

Rontagen	Patologi Anatomis
efusi pleura kiri	sitologi : SCC
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral	sputum : BTA +++
efusi pleura unilateral	sputum : BTA ++, segmen posterior lobus superior
efusi pleura unilateral, infiltrat	sputum : BTA +++
efusi pleura unilateral	sputum : BTA +++
efusi pleura	sitologi : SCC
efusi pleura	sitologi : SCC
efusi pleura kanan	sitologi : SCC
efusi pleura kanan	sitologi : NSCC
efusi pleura kanan	sitologi : SCC
efusi pleura bilateral	sitologi : NSCC
efusi pleura kanan	sitologi : SCC
efusi pleura kanan	sitologi : SCC
efusi pleura kiri massif	sitologi : NSCC
efusi pleura kanan massif	sitologi : NSCC
effusi pleura	sitologi : SCC
efusi pleura massif kanan	sitologi : NSCC
efusi pleura massif kanan	sitologi : NSCC
efusi pleura massif, metastasis intrapulmoner	sitologi : NSCC
efusi pleura unilateral, infiltrat	sputum : BTA ++, segmen posterior lobus superior
efusi pleura unilateral, infiltrat, kaviti dengan dinding tebal	sputum : BTA ++, segmen posterior lobus superior
efusi pleura unilateral	sputum : BTA +++
efusi pleura	sitologi : SCC
efusi pleura kiri massif	sitologi : NSCC

Rontagen	Patologi Anatomis
efusi pleura kanan massif	sitologi : NSCC
efusi pleura non massif	sitologi : SCC
efusi pleura bilateral	sitologi : NSCC
efusi pleura kanan massif	sitologi : NSCC
efusi pleura	sitologi : SCC
efusi pleura	sitologi : SCC
efusi pleura	sitologi : SCC
efusi pleura kanan massif	sitologi : NSCC
efusi pleura kanan massif	sitologi : NSCC
efusi pleura	sitologi : SCC

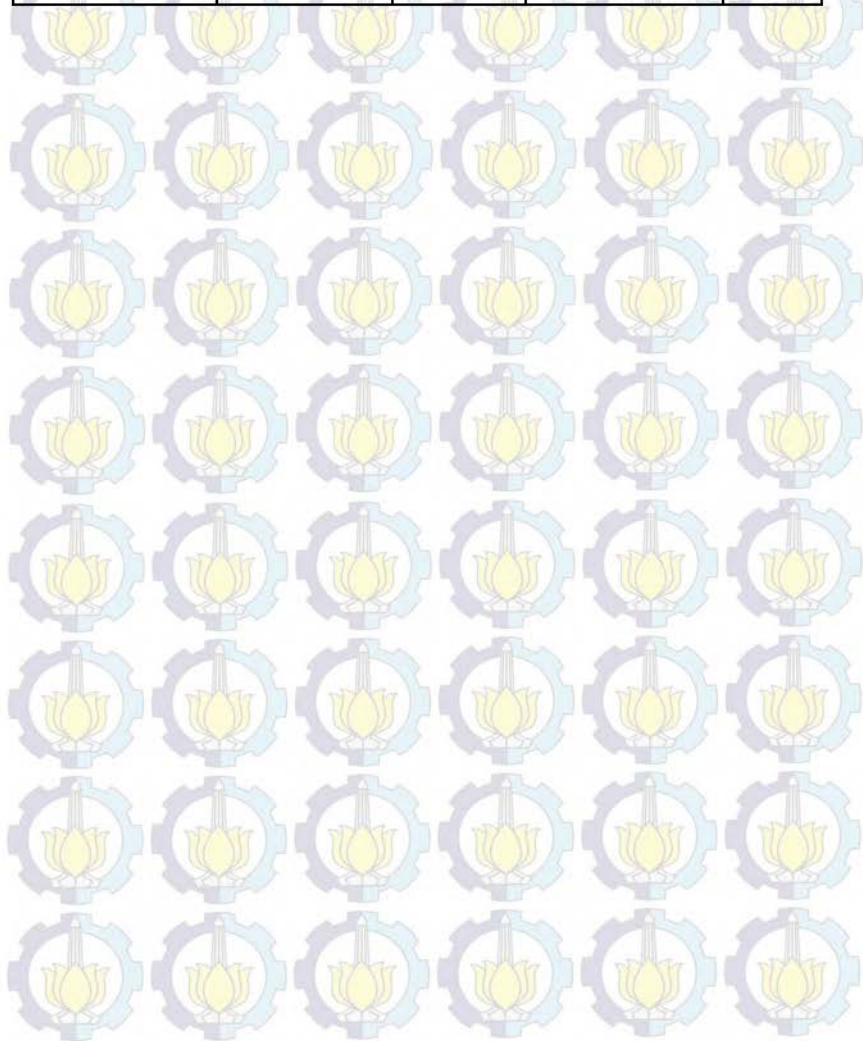
Lanjutan tabel :

Positif kanker paru	batuk berdahak	sesak nafas	nyeri daerah dada	dema m
ya	ya	ya	tidak	tidak
ya	ya	tidak	ya	tidak
ya	ya	tidak	ya	ya
ya	tidak	ya	ya	tidak
tidak	tidak	ya	tidak	tidak
tidak	ya	tidak	ya	ya
tidak	tidak	ya	ya	tidak
ya	tidak	tidak	ya	tidak
tidak	tidak	ya	ya	tidak
tidak	tidak	tidak	ya	tidak
tidak	tidak	ya	tidak	tidak
ya	ya	ya	tidak	tidak
ya	tidak	tidak	tidak	tidak
ya	ya	ya	ya	tidak
tidak	tidak	tidak	tidak	ya
tidak	ya	tidak	ya	tidak
tidak	ya	tidak	ya	tidak
tidak	tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak	tidak
tidak	tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak	tidak
ya	ya	ya	tidak	ya
ya	ya	ya	tidak	tidak
ya	ya	ya	tidak	ya
ya	ya	ya	ya	tidak

Positif kanker paru	batuk berdahak	sesak nafas	nyeri daerah dada	dema m
ya	tidak	tidak	tidak	tidak
ya	ya	ya	ya	tidak
ya	ya	ya	ya	tidak
ya	ya	ya	tidak	ya
ya	ya	ya	tidak	tidak
tidak	tidak	ya	tidak	tidak
tidak	tidak	ya	tidak	tidak
tidak	tidak	tidak	tidak	tidak
tidak	ya	ya	tidak	tidak
tidak	tidak	ya	ya	tidak
tidak	ya	tidak	tidak	tidak
tidak	tidak	tidak	ya	ya
tidak	ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak	ya
ya	tidak	ya	tidak	tidak
ya	ya	ya	tidak	tidak
ya	tidak	ya	tidak	tidak
ya	ya	ya	tidak	tidak
ya	tidak	ya	ya	tidak
ya	ya	tidak	ya	ya
ya	ya	ya	ya	tidak
tidak	ya	tidak	tidak	ya
tidak	tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak	ya
tidak	tidak	ya	ya	tidak
ya	tidak	tidak	ya	tidak

Positif kanker paru	batuk berdahak	sesak nafas	nyeri daerah dada	dema m
ya	ya	tidak	ya	tidak
ya	tidak	ya	ya	tidak
ya	tidak	tidak	tidak	tidak
ya	ya	ya	tidak	tidak
ya	tidak	ya	ya	tidak
ya	ya	tidak	ya	tidak
ya	ya	tidak	ya	tidak
ya	tidak	ya	tidak	tidak
ya	tidak	ya	ya	ya
ya	ya	ya	tidak	tidak
ya	tidak	tidak	tidak	ya
ya	tidak	tidak	ya	tidak
ya	tidak	tidak	tidak	ya
tidak	ya	tidak	tidak	tidak
tidak	ya	tidak	tidak	tidak
tidak	tidak	tidak	ya	tidak
ya	ya	tidak	tidak	tidak
ya	ya	tidak	tidak	tidak
ya	tidak	tidak	tidak	tidak
ya	tidak	ya	ya	tidak
ya	ya	ya	tidak	tidak
ya	tidak	tidak	ya	tidak
ya	ya	ya	tidak	ya
ya	ya	ya	tidak	tidak
ya	tidak	tidak	ya	tidak
ya	ya	tidak	ya	tidak
ya	ya	ya	tidak	ya
ya	ya	ya	tidak	tidak
ya	tidak	tidak	ya	tidak
ya	ya	tidak	ya	tidak
ya	ya	ya	tidak	tidak

Positif kanker paru	batuk berdahak	sesak nafas	nyeri daerah dada	dema m
ya	ya	tidak	ya	tidak



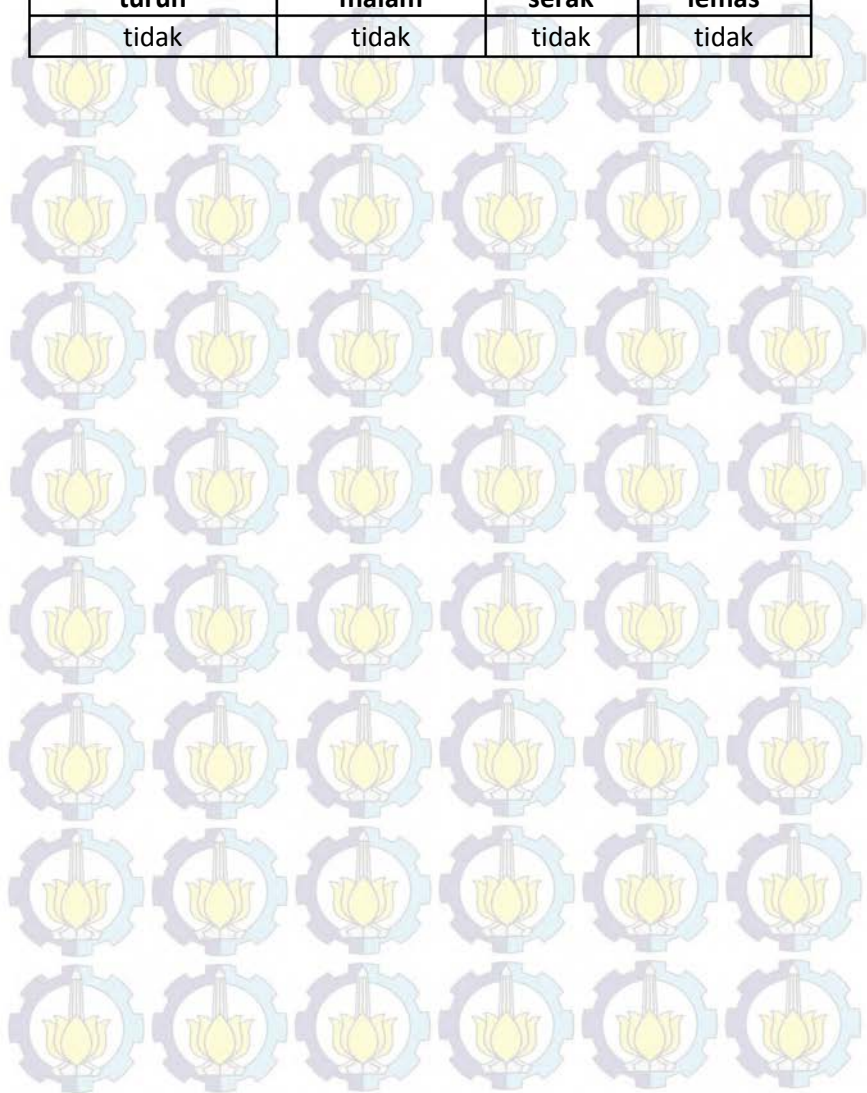
Lanjutan tabel :

nafsu makan turun	keringat malam	suara serak	lelah / lemas
Ya	tidak	tidak	tidak
Tidak	tidak	tidak	ya
Tidak	tidak	tidak	tidak
Tidak	tidak	tidak	tidak
Tidak	tidak	tidak	tidak
Ya	tidak	tidak	tidak
Tidak	tidak	tidak	tidak
Tidak	tidak	tidak	ya
Tidak	tidak	tidak	tidak
Ya	ya	tidak	tidak
Tidak	ya	tidak	tidak
Tidak	tidak	tidak	tidak
Tidak	tidak	tidak	tidak
Tidak	tidak	tidak	tidak
Tidak	ya	ya	tidak
Tidak	tidak	tidak	tidak
Tidak	tidak	tidak	tidak
Tidak	ya	ya	tidak
Tidak	tidak	tidak	ya
Ya	ya	ya	tidak
Tidak	tidak	tidak	ya
Tidak	tidak	tidak	tidak
Tidak	tidak	tidak	ya
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak

nafsu makan turun	keringat malam	suara serak	lelah / lemas
tidak	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	ya
tidak	ya	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	ya
tidak	tidak	tidak	ya
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak

nafsu makan turun	keringat malam	suara serak	lelah / lemas
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	ya	tidak	tidak
tidak	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	ya	tidak	tidak
ya	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak
tidak	ya	tidak	tidak
tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak

nafsu makan turun	keringat malam	suara serak	lelah / lemas
tidak	tidak	tidak	tidak



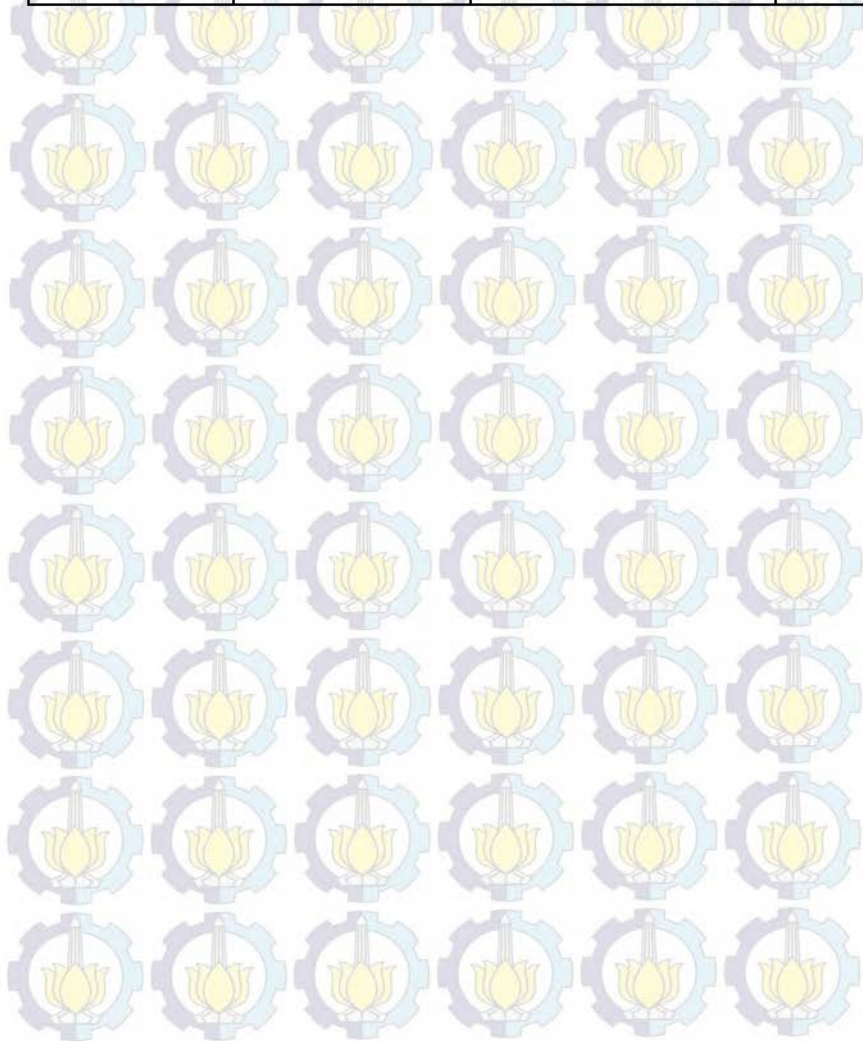
Lanjutan tabel:

batuk berdarah	berat badan turun	benjolan pangkal leher	pusing
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	ya	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak

batuk berdarah	berat badan turun	benjolan pangkal leher	pusing
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	ya	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	ya	tidak	tidak
tidak	ya	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	ya
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	ya	tidak	tidak
tidak	tidak	tidak	tidak
ya	ya	tidak	tidak
ya	tidak	tidak	tidak

batuk berdarah	berat badan turun	benjolan pangkal leher	pusing
tidak	ya	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	tidak	tidak
ya	tidak	ya	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak
ya	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	ya	tidak
tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak
tidak	tidak	ya	tidak
tidak	tidak	tidak	tidak
tidak	tidak	tidak	tidak

batuk berdarah	berat badan turun	benjolan pangkal leher	pusing
tidak	tidak	tidak	tidak



C. Listing Program

1. Preprocessing.m

```

clear all;
clc;
time_start=tic;
a1=xlsread('book5.xlsx','A2:A81');
a2=xlsread('book5.xlsx','B2:B81');
a3=xlsread('book5.xlsx','C2:C81');
a4=xlsread('book5.xlsx','D2:D81');
a5=xlsread('book5.xlsx','E2:E81');
a6=xlsread('book5.xlsx','F2:F81');

m=length(a1);
for i1=1:m
    if a1(i1)==1
        class(i1)={'Kanker Paru'};
    else
        class(i1)={'Bukan Kanker Paru'};
    end
end

for i1=1:m
    if a2(i1)==1
        b1(i1)={'Batuk berdahak'};
    else
        b1(i1)={'Bukan Batuk berdahak'};
    end
end

for i1=1:m
    if a3(i1)==1
        b2(i1)={'Sesak'};
    else
        b2(i1)={'Bukan sesak'};
    end
end

```

```

end

for i1=1:m
    if a4(i1)==1
        b3(i1)={'Nyeri Daerah Dada'};
    else
        b3(i1)={'Bukan Nyeri Daerah Dada'};
    end
end

for i1=1:m
    if a5(i1)==1
        b4(i1)={'Batuk Berdarah'};
    else
        b4(i1)={'Bukan Batuk Berdarah'};
    end
end

for i1=1:m
    if a6(i1)==1
        b5(i1)={'Sitologi Klinis'};
    else
        b5(i1)={'Suptum'};
    end
end

atribut=[b1' b2' b3' b4' b5' ];
save atribut1 atribut;
save kelas class;
time_end=toc(time_start)

```

2. CBS.m

```

clear all;
clc;
time_start=tic;
load kelas;
load atribut1;
Dt=[atribut class'];
jum_train=0.8*length(class);
%algorithm CBS
%CSP Miner algorithm
minsup=input('minimum support yang diberikan
adalah ');
class=unique(Dt(:,6));
x=0;k=0;n=1;pp=0;
for i1=1:size(class,1)
    for j1=1:jum_train
        if
isequal(Dt(j1,6),class(i1,1))==1
            x=x+1;
            Dc1(x,:)=Dt(j1,:);
        end
    end
    [N,M]=size(Dc1);
    min_sup=minsup*N;
    x=x-x;
    %find the frequent itemset F1
    f_candidate(:,1)=unique(Dc1(:,1:5));
    for j2=1:size(f_candidate,1)
        fre1(j2)=0;
        f1(j2)=0;
        for i4=1:N
            if
sum(ismember(Dc1(i4,:),f_candidate(j2,1)))>0
                fre1(j2)=fre1(j2)+1;
            else
                fre1(j2)=fre1(j2);
            end
        end
    end
end

```

```

        if fre1(j2)>=min_sup
            k=k+1;
            x=x+1;
            f1(k)=fre1(j2);
            F1(x)=f_candidate(j2);
            T1(k,1)=F1(x);
            CSP_SET(k,1)=F1(x);
            indeks_csp(k)=class(i1);
        end
    end
    x=x-x;
    fre1=[];
    f1=[];
    %make CSP SET
    while n+1<=6
        n=n+1;
        for p1=1:size(T1)
            for q1=1:size(F1')
                x=x+1;

Textension1(x,1:n)=[T1(p1,1:n-1) F1(q1)];
            end
        end
        x=x-x;t2=1;jj1=0;T1={};
        for p2=1:size(Textension1,1)
            fr1(p2)=0;
            for jj=1:N
                while
t2<=size(Textension1',1)&jj1+1<=M
                    jj1=jj1+1;
                    if
isequal(Dc1(jj,jj1),Textension1(p2,t2))==1
                        t2=t2+1;
                    end
                end
            end
            if t2==size(Textension1',1)+1
                fr1(p2)=fr1(p2)+1;
            else
                fr1(p2)=fr1(p2);
            end
        end
    end
end

```



```

        jj1=jj1-jj1;
        t2=t2-t2+1;
    end
    if fr1(p2)>=min_sup
        k=k+1;x=x+1;pp=pp+1;
        fel(x)=fr1(p2);
        T1(x,1:n)=Textension1(p2,1:n);

CSP_SET(k,1:n)=Textension1(p2,1:n);
        indekscsp(k)=class(i1);
    end
end
fr1=[];
fel=[];
x=x-x;
Textension1={};
end
n=1;
N=N-N;
M=M-M;
Dc1={};
f_candidate={};
Textension1={};
F1={};
T1={};
min_sup=[];
end
k=k-k;
csp_set(:,1:size(CSP_SET',1))=CSP_SET;
csp_set(:,size(CSP_SET',1)+1)={'=>'};
csp_set(:,size(CSP_SET',1)+2)=indekscsp;
% Classifier Builder algorithm
for i3=1:size(CSP_SET,1)%make csp_dup
    count(i3)=0;
    for i2=1:size(CSP_SET,1)
        if
isequal(CSP_SET(i3,:),CSP_SET(i2,:))==1
            count(i3)=count(i3)+1;
        end
    end
end
end

```

```

        if count(i3)==size(class,1)
            k=k+1;
            csp_dup(k,:)=CSP_SET(i3,:);
        end
    end
    k=k-k;
    a(1,1:size(CSP_SET',1))={ ' ' };
    h=exist('csp_dup');
    if h==1
        for i5=1:size(csp_dup)
            for j3=1:size(CSP_SET)
                if
                    isequal(csp_dup(i5,:),CSP_SET(j3,:))==1
                        CSP_SET(j3,:)=a;
                    end
                end
            end
        for j3=1:size(CSP_SET,1)
            if isequal(CSP_SET(j3,:),a)==0
                k=k+1;

                T(k,1:size(CSP_SET',1))=CSP_SET(j3,:);
                T(k,size(CSP_SET',1)+1)={'=>'};

                T(k,size(CSP_SET',1)+2)=indekscsp(j3);
            end
        end
        T
        save cspmodif T;
    else
        T=csp_set
        save cspmodif T;
    end
    time_end=toc(time_start);

```

3. Akurasi.m

```

clear all;
clc;
%count accuracy
load kelas;
load atribut1;
load cspmodif;
jum_train=0.2*length(class);
D=[atribut class'];
for p=1:jum_train
    Dtes(p,:)=D(p,:);
    Dbodi(p,:)=D(p,:);
end
x=0;countbodi=0;
counttes=0;
for k=1:size(T,1)
    for kk=1:size(T',1)-3
        if isequal(T(k,kk),{[]})==0
            t=T(k,1:kk);
            for jj=1:jum_train
                if
sum(ismember(Dtes(jj,:),t))>0&isequal(T(k,size
e(T',1)),Dtes(jj,size(Dtes',1)))==1
                    counttes=counttes+1;
                    Dtes(jj,:)='{ ' '};
                end
            if
sum(ismember(Dbodi(jj,:),t))>0
                    countbodi=countbodi+1;
                    Dbodi(jj,:)='{ ' '};
                end
            end
        end
    end
    t={};
end
coverage=(countbodi/jum_train)*100
accuracy=(counttes/countbodi)*100
covacc=(accuracy/100)*(1+(coverage/100))

```

4. Interface.m

```

function varargout = TA(varargin)
% TA M-file for TA.fig
%     TA, by itself, creates a new TA or
%     raises the existing
%     singleton*.
%
%     H = TA returns the handle to a new TA
%     or the handle to
%     the existing singleton*.
%
%
TA('CALLBACK', hObject,eventData,handles,...)
calls the local
%     function named CALLBACK in TA.M with
%     the given input arguments.
%
%     TA('Property','Value',...) creates a
%     new TA or raises the
%     existing singleton*. Starting from
%     the left, property value pairs are
%     applied to the GUI before
TA_OpeningFcn gets called. An
%     unrecognized property name or invalid
%     value makes property application
%     stop. All inputs are passed to
TA_OpeningFcn via varargin.
%
%     *See GUI Options on GUIDE's Tools
%     menu. Choose "GUI allows only one
%     instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response
% to help TA

```

```

% Last Modified by GUIDE v2.5 16-Jan-2015
00:33:47

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',
mfilename, ...
                  'gui_Singleton',
gui_Singleton, ...
                  'gui_OpeningFcn',
@TA_OpeningFcn, ...
                  'gui_OutputFcn',
@TA_OutputFcn, ...
                  'gui_LayoutFcn', [], ...
                  'gui_Callback', []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback =
str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] =
gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT

% --- Executes just before TA is made
visible.
function TA_OpeningFcn(hObject, eventdata,
handles, varargin)
% This function has no output args, see
OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a
future version of MATLAB

```

```

% handles      structure with handles and user
data (see GUIDATA)
% varargin     command line arguments to TA
               (see VARARGIN)

% Choose default command line output for TA
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% UIWAIT makes TA wait for user response (see
UIRESUME)
% uiwait(handles.figure1);

set(handles.edit3,'Visible','off');
set(handles.text3,'Visible','off');
set(handles.edit5,'Visible','off');
set(handles.text8,'Visible','off');

% --- Outputs from this function are returned
to the command line.
function varargout = TA_OutputFcn(hObject,
eventdata, handles)
% varargout    cell array for returning output
args (see VARARGOUT);
% hObject     handle to figure
% eventdata   reserved - to be defined in a
future version of MATLAB
% handles     structure with handles and user
data (see GUIDATA)

% Get default command line output from
handles structure
varargout{1} = handles.output;

```

```

function edit1_Callback(hObject, eventdata,
handles)
% hObject      handle to edit1 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      structure with handles and user
data (see GUIDATA)

% Hints: get(hObject,'String') returns
contents of edit1 as text
%           str2double(get(hObject,'String'))
returns contents of edit1 as a double

% --- Executes during object creation, after
setting all properties.
function edit1_CreateFcn(hObject, eventdata,
handles)
% hObject      handle to edit1 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      empty - handles not created
until after all CreateFcns called

% Hint: edit controls usually have a white
background on Windows.
%           See ISPC and COMPUTER.
if ispc &&
isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

```

```

% --- Executes on button press in browse.
function browse_Callback(hObject, eventdata, handles)
% hObject      handle to browse (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      structure with handles and user
data (see GUIDATA)
%Membuka kotak dialog untuk mengambil file,
sebagai variabel filename
filename = uigetfile('*.xlsx');

%Digunakan untuk mencegah error jk user tdk
jd mengambil file
if isequal(filename,0)
    clear all;
    clc
    return
end

%Menampilkan nama file pada text1
set(handles.edit1,'String',filename);
%Membaca file excel ,&mnyimpannya sbg
variabel data
preprop
% --- Executes on button press in
pushbutton2.
function pushbutton2_Callback(hObject,
eventdata, handles)
% hObject      handle to pushbutton2 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      structure with handles and user
data (see GUIDATA)

function edit2_Callback(hObject, eventdata,
handles)

```



```
% hObject      handle to edit2 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      structure with handles and user
data (see GUIDATA)
```

```
% Hints: get(hObject,'String') returns
contents of edit2 as text
%          str2double(get(hObject,'String'))
returns contents of edit2 as a double
```

```
% --- Executes during object creation, after
setting all properties.
```

```
function edit2_CreateFcn(hObject, eventdata,
handles)
```

```
% hObject      handle to edit2 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      empty - handles not created
until after all CreateFcns called
```

```
% Hint: edit controls usually have a white
background on Windows.
```

```
%          See ISPC and COMPUTER.
```

```
if ispc &&
isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUiControlBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

```
function edit3_Callback(hObject, eventdata,
handles)
```

```
% hObject      handle to edit3 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
```

```

% handles      structure with handles and user
data (see GUIDATA)

% Hints: get(hObject,'String') returns
contents of edit3 as text
%          str2double(get(hObject,'String'))
returns contents of edit3 as a double

% --- Executes during object creation, after
setting all properties.
function edit3_CreateFcn(hObject, eventdata,
handles)
% hObject      handle to edit3 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      empty - handles not created
until after all CreateFcns called

% Hint: edit controls usually have a white
background on Windows.
%          See ISPC and COMPUTER.
if ispc &&
isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function edit4_Callback(hObject, eventdata,
handles)
% hObject      handle to edit4 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      structure with handles and user
data (see GUIDATA)
global A;

```

```

A=str2double(get(handles.edit4, 'string'));
% Hints: get(hObject,'String') returns
contents of edit4 as text
%         str2double(get(hObject,'String'))
returns contents of edit4 as a double

% --- Executes during object creation, after
setting all properties.
function edit4_CreateFcn(hObject, eventdata,
handles)
% hObject    handle to edit4 (see GCBO)
% eventdata  reserved - to be defined in a
future version of MATLAB
% handles    empty - handles not created
until after all CreateFcns called

% Hint: edit controls usually have a white
background on Windows.
%         See ISPC and COMPUTER.
if ispc &&
isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Executes on button press in startpre.
function startpre_Callback(hObject,
eventdata, handles)
% hObject    handle to startpre (see GCBO)
% eventdata  reserved - to be defined in a
future version of MATLAB
% handles    structure with handles and user
data (see GUIDATA)
global class;
global A;
h = msgbox('Preprocessing Complete');
load kelas;

```

```

load atribut1;
Dt=[atribut class'];
jum_train=0.8*length(class);
%algorithm CBS
%CSP Miner algorithm
minsup=A
class=unique(Dt(:,6));
x=0;k=0;n=1;pp=0;
for i1=1:size(class,1)
    for j1=1:jum_train
        if
isequal(Dt(j1,6),class(i1,1))==1
            x=x+1;
            Dc1(x,:)=Dt(j1,:);
        end
    end
    [N,M]=size(Dc1);
    min_sup=minsup*N;
    x=x-x;
    %find the frequent itemset F1
    f_candidate(:,1)=unique(Dc1(:,1:5));
    for j2=1:size(f_candidate,1)
        fre1(j2)=0;
        f1(j2)=0;
        for i4=1:N
            if
sum(ismember(Dc1(i4,:),f_candidate(j2,1)))>0
                fre1(j2)=fre1(j2)+1;
            else
                fre1(j2)=fre1(j2);
            end
        end
    end
    if fre1(j2)>=min_sup
        k=k+1;
        x=x+1;
        f1(k)=fre1(j2);
        F1(x)=f_candidate(j2);
        T1(k,1)=F1(x);
        CSP_SET(k,1)=F1(x);
        indeks_csp(k)=class(i1);
    end
end

```

```

        end
    end
    x=x-x;
    fre1=[];
    f1=[];
    %make CSP SET
    while n+1<=13
        n=n+1;
        for p1=1:size(T1)
            for q1=1:size(F1')
                x=x+1;

Textension1(x,1:n)=[T1(p1,1:n-1) F1(q1)];
            end
        end
        x=x-x;t2=1;jj1=0;T1={};
        for p2=1:size(Textension1,1)
            fr1(p2)=0;
            for jj=1:N
                while
t2<=size(Textension1',1)&jj1+1<=M
                    jj1=jj1+1;
                    if
isequal(Dc1(jj,jj1),Textension1(p2,t2))==1
                        t2=t2+1;
                    end
                end
            end
            if t2==size(Textension1',1)+1
                fr1(p2)=fr1(p2)+1;
            else
                fr1(p2)=fr1(p2);
            end
            jj1=jj1-jj1;
            t2=t2-t2+1;
        end
    end
    if fr1(p2)>=min_sup
        k=k+1;x=x+1;pp=pp+1;
        fel(x)=fr1(p2);
        T1(x,1:n)=Textension1(p2,1:n);
    end
end

```

```

CSP_SET(k,1:n)=Textension1(p2,1:n);
    indekscsp(k)=class(i1);
    end
    end
    fr1=[];
    fe1=[];
    x=x-x;
    Textension1={};
    end
    n=n-n+1;
    N=N-N;
    M=M-M;
    Dc1={};
    f_candidate={};
    Textension1={};
    F1={};
    T1={};
    min_sup=[];
end
k=k-k;
csp_set(:,1:size(CSP_SET',1))=CSP_SET;
csp_set(:,size(CSP_SET',1)+1)={'=>'};
csp_set(:,size(CSP_SET',1)+2)=indekscsp;
% Classifier Builder algorithm
h = msgbox('Operation CBS');
for i3=1:size(CSP_SET,1)%make csp_dup
    count(i3)=0;
    for i2=1:size(CSP_SET,1)
        if
            isequal(CSP_SET(i3,:),CSP_SET(i2,:))==1
                count(i3)=count(i3)+1;
            end
        end
        if count(i3)==size(class,1)
            k=k+1;
            csp_dup(k,:)=CSP_SET(i3,:);
        end
    end
end
k=k-k;

```

```

a(1,1:size(CSP_SET',1))={' '};
h=exist('csp_dup');
if h==1
for i5=1:size(csp_dup)
    for j3=1:size(CSP_SET)
        if
isequal(csp_dup(i5,:),CSP_SET(j3,:))==1
            CSP_SET(j3,:)=a;
        end
    end
end
for j3=1:size(CSP_SET,1)
    if isequal(CSP_SET(j3,:),a)==0
        k=k+1;

T(k,1:size(CSP_SET',1))=CSP_SET(j3,:);
T(k,size(CSP_SET',1)+1)={'=>'};

T(k,size(CSP_SET',1)+2)=indekscsp(j3);
    end
end
T
save cspmodif T;
else
    T=csp_set
    save cspmodif T;
end

set(handles.edit3,'Visible','on');
set(handles.text3,'Visible','on');
set(handles.edit5,'Visible','on');
set(handles.text8,'Visible','on');

load kelas;
load atribut1;
load cspmodif;
jum_train=0.2*length(class);
D=[atribut class'];

```

```

for p=1:jum_train
    Dtes(p,:)=D(p,:);
    Dbodi(p,:)=D(p,:);
end
x=0;countbodi=0;
counttes=0;
for k=1:size(T,1)
    for kk=1:size(T',1)-3
        if isequal(T(k,kk),{[]})==0
            t=T(k,1:kk);
            for jj=1:jum_train
                if
sum(ismember(Dtes(jj,:),t))>0&isequal(T(k,size(T',1)),Dtes(jj,size(Dtes',1)))==1
                    counttes=counttes+1;
                    Dtes(jj,:)={' '};
                end
                if
sum(ismember(Dbodi(jj,:),t))>0
                    countbodi=countbodi+1;
                    Dbodi(jj,:)={' '};
                end
            end
        end
    end
    t={};
end
h = msgbox('Operation Completed');
coverage=(countbodi/jum_train)*100
accuracy=(counttes/countbodi)*100
covacc=(accuracy/100)*(1+(coverage/100))
set(handles.edit3,'String',accuracy);
set(handles.edit5,'String',coverage);

function edit5_Callback(hObject, eventdata,
handles)

```



```
% hObject      handle to edit5 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      structure with handles and user
data (see GUIDATA)
```

```
% Hints: get(hObject,'String') returns
contents of edit5 as text
%          str2double(get(hObject,'String'))
returns contents of edit5 as a double
```

```
% --- Executes during object creation, after
setting all properties.
```

```
function edit5_CreateFcn(hObject, eventdata,
handles)
```

```
% hObject      handle to edit5 (see GCBO)
% eventdata    reserved - to be defined in a
future version of MATLAB
% handles      empty - handles not created
until after all CreateFcns called
```

```
% Hint: edit controls usually have a white
background on Windows.
```

```
%          See ISPC and COMPUTER.
```

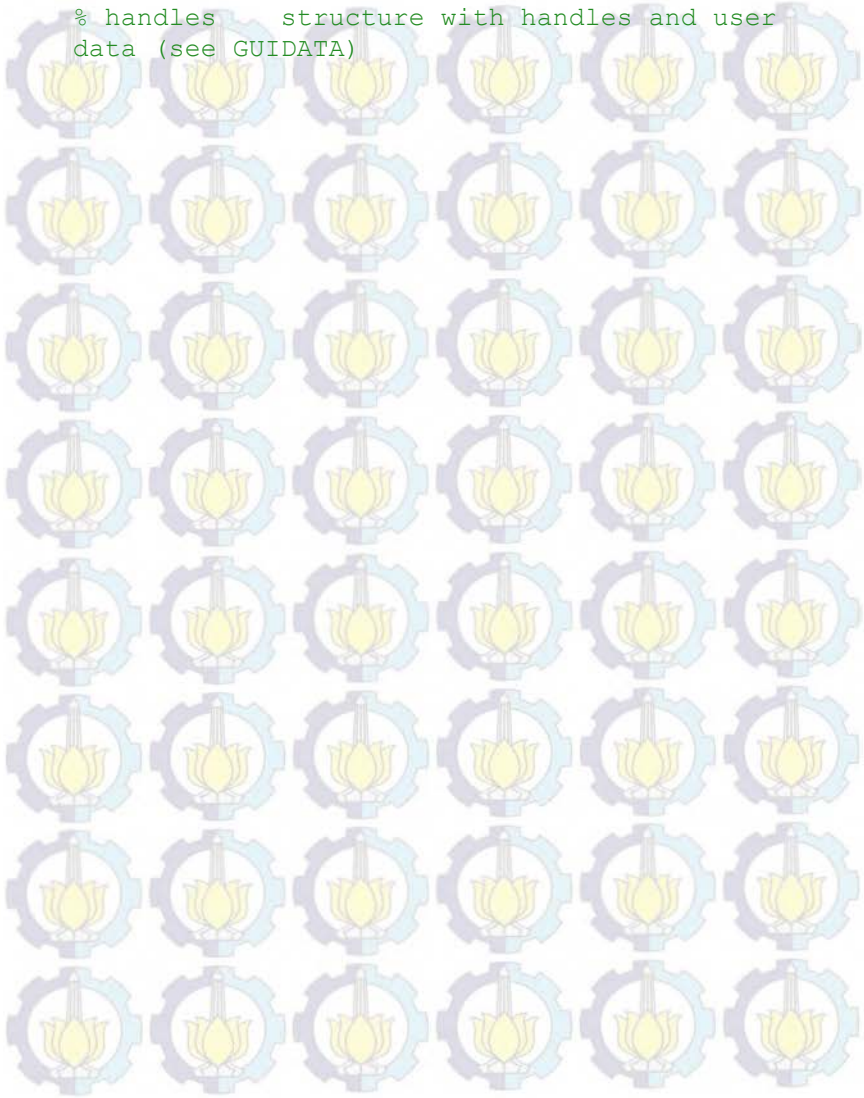
```
if ispc &&
isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

```
% -----
-----
```

```
function Untitled_1_Callback(hObject,
eventdata, handles)
```

```
% hObject      handle to Untitled_1 (see GCBO)
```

```
% eventdata reserved - to be defined in a  
future version of MATLAB  
% handles structure with handles and user  
data (see GUIDATA)
```



D. Hasil Running

- Min_sup 0.9



Gambar D1. Min Sup 0,9

- Min_sup 0.8



Gambar D2. Min Sup 0,8

- Min_sup 0.7



Gambar D3. Min Sup 0,7

- Min_sup 0.6



Gambar D4. Min Sup 0,6

- Min_sup 0.5

TA

Untitled 1

**KLASIFIKASI KANKER PARU DENGAN
MENGGUNAKAN ALGORITMA
CLASSIFY BY SEQUENCE**

Data Set

Book3.xlsx

Browse...

Accuracy Level (%)

100

Minimum Support

0.5

Coverage (%)

100

Start Preprocessing

I Putu Eddy Sukadiana
1206100711

Gambar D5. Min Sup 0,5

- Min_sup 0.4



Gambar D6. Min Sup 0,4

- Min_sup 0.3



Gambar D7. Min Sup 0,3

- Min_sup 0.2



Gambar D8. Min Sup 0,2

- Min_sup 0.1



Gambar D9. Min Sup 0,1

E. Tabel Hasil Running Beserta Waktunya

No	<i>Min_sup</i>	Accuracy	Coverage	Waktu
1	0.9	-	-	4.0125066
2	0.8	-	0	4.47270743
3	0.7	100	100	2.78582629
4	0.6	100	100	4.70338988
5	0.5	100	100	8.34980304
6	0.4	100	100	10.6992991
7	0.3	100	100	18.8501695
8	0.2	100	100	49.7701854
9	0.1	100	100	1.23E+02

**F. Tabel data yang digunakan untuk penjalanan program
(setelah dilakukan data *selection* dan data *transformation*)**

Positif kanker paru	batuk berdahak	sesak nafas
1	1	1
1	1	0
0	1	0
0	0	1
1	0	0
1	1	1
0	1	0
1	1	1
0	0	0
1	1	1

Positif kanker paru	batuk berdahak	sesak nafas
0	0	1
1	1	1
1	0	0
1	1	1
1	1	1
0	1	1
1	1	1
1	0	1
1	1	1
1	0	1
1	1	1
1	0	1
1	1	0
1	1	1
1	0	0
0	0	0
1	0	1
1	0	0
1	1	1
1	0	1
0	1	0
1	1	0
1	0	1
1	0	1
1	1	1
1	0	0
1	0	0

Positif kanker paru	batuk berdahak	sesak nafas
1	0	0
1	1	0
1	1	0
0	0	0
1	0	1
1	1	1
1	0	0
1	1	1
1	1	1
1	0	0
1	1	0
1	1	1
1	1	0
0	0	1
0	1	0
0	0	1
0	0	1
1	1	1
0	0	1
0	0	0
0	1	0
0	1	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	1

Positif kanker paru	batuk berdahak	sesak nafas
1	1	1
0	0	0
1	1	0
1	1	0
0	1	0
0	0	0
1	1	0
0	0	0
1	0	1
1	0	0
0	0	0
1	1	1
0	0	1
1	0	0
0	1	0
0	0	0

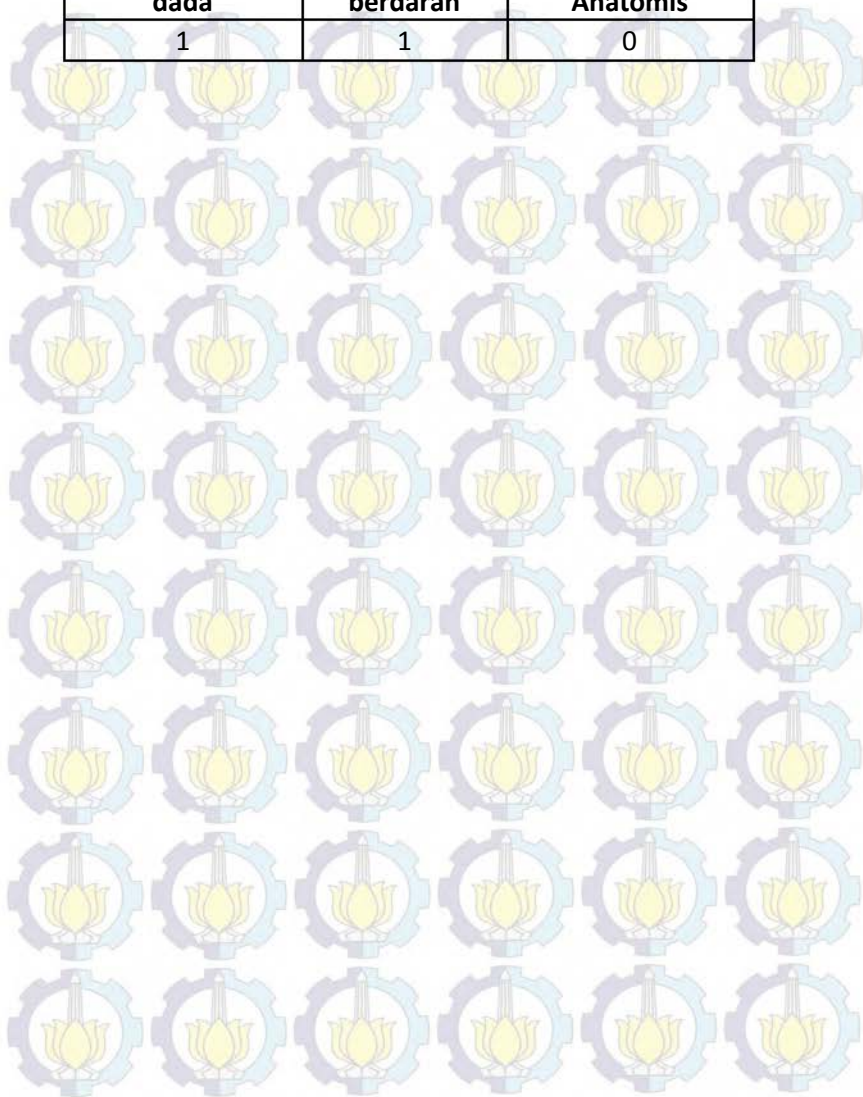
Lanjutan tabel :

nyeri daerah dada	batuk berdarah	Patologi Anatomis
0	0	1
1	0	1
0	0	0
0	1	0
1	0	1
0	0	1
0	0	0
1	0	1
0	0	0
0	0	1
1	1	0
1	0	1
0	1	1
1	0	1
1	0	1
0	1	0
0	0	1
0	1	1
0	1	1
0	1	1
0	0	1
1	0	1
1	0	1
1	0	1
1	1	1

nyeri daerah dada	batuk berdarah	Patologi Anatomis
1	0	0
1	0	1
0	1	1
0	0	1
1	1	1
0	0	0
1	0	1
0	0	1
1	0	1
0	0	1
0	1	1
1	1	1
0	1	1
0	0	1
0	0	1
0	1	0
1	0	1
0	0	1
1	0	1
0	0	1
0	0	1
1	0	1
1	0	1
0	0	1
0	0	1
1	0	1
1	0	1
0	1	0
1	0	0

nyeri daerah dada	batuk berdarah	Patologi Anatomis
1	0	0
1	0	0
0	0	1
0	1	0
0	1	0
1	0	0
1	1	0
1	1	0
0	1	0
1	1	0
0	1	0
0	0	0
0	0	1
0	1	0
1	0	1
1	0	1
0	1	0
1	0	0
1	0	1
0	1	0
1	0	1
0	1	1
0	1	0
0	0	1
1	1	0
0	1	1
0	0	0

nyeri daerah dada	batuk berdarah	Patologi Anatomis
1	1	0



"Halaman ini sengaja dikosongkan"

BIODATA PENULIS



Penulis memiliki nama lengkap I Putu Eddy Sukadiana atau biasa dipanggil Eddy. Penulis dilahirkan di Denpasar pada 27 Februari 1987 dan merupakan anak pertama dari 2 bersaudara. Pendidikan formal yang pernah ditempuh yaitu di TK Astiti Denpasar, SDN 17 Dauh Puri Denpasar, SMPN 5 Denpasar, dan SMKN 1 Denpasar. Setelah lulus dari SMK, penulis diterima di Jurusan Matematika FMIPA ITS pada tahun 2006 dan terdaftar

dengan NRP. 1206 100 711.

Di Jurusan Matematika FMIPA ITS, penulis mengambil bidang minat Ilmu Komputer. Penulis aktif di beberapa organisasi intra kampus diantaranya kepengurusan HIMATIKA sebagai staf Departemen Pendidikan periode 2007-2008 dan periode 2008-2009. Penulis juga aktif dalam berbagai kegiatan pelatihan kemahasiswaan, diantaranya LKMM Pra TD FMIPA 2007, LKMM TD HIMATIKA 2008. Untuk kritik, saran, informasi dan pertanyaan yang berhubungan mengenai Tugas Akhir ini dapat dikirimkan melalui *e-mail* ke eddy.sukadiana@gmail.com.